

Diseño y realización de proyectos de investigación sobre sistemas de salud, *volumen II,*

Análisis de los datos y redacción del informe

Corlien M. Varkevisser
Indra Pathmanathan
Ann Brownlee

Diseño y realización de proyectos de investigación sobre sistemas de salud

Volumen II: análisis de los datos y redacción del informe

Este espacio en blanco intencionalmente dejado de la página

Diseño y realización de proyectos de investigación sobre sistemas de salud

Volumen II: análisis de los datos y redacción del informe

*Corlien M. Varkevisser
Indra Pathmanathan
Ann Brownlee*

Centro Internacional de Investigaciones para el Desarrollo

Ottawa • Dakar • El Cairo • Montevideo • Nairobi • Nueva Delhi • Singapur



© Centro Internacional de Investigaciones para el Desarrollo 2011

Diseño y realización de proyectos de investigación sobre sistemas de salud
Volumen II: Análisis de los datos y redacción del informe

Primera edición: julio de 2011

Copublicado en castellano por:

Mayol Ediciones S.A.

Bogotá, Colombia

www.mayolediciones.com

ISBN: 978-958-8307-95-4

Centro Internacional de Investigaciones para el Desarrollo

PO Box 8500

Ottawa, ON, Canadá K1G 3H9

info@idrc.ca

www.idrc.ca

ISBN: 978-1-55250-512-0

Todos los derechos reservados

Traducción al castellano: IDRC

Diseño de cubierta: Mulder en van Meurs, Amsterdam

Edición y diagramación: Mayol Ediciones S.A.

Impreso en Colombia - Printed in Colombia

CONTENIDO

Prólogo		vii
Prefacio y reconocimientos		ix
Introducción a la parte II: Análisis de los datos y redacción del informe		xiii
Módulo 21	Orientación para el taller de análisis de los datos y redacción del informe	3
Módulo 22	Descripción de las variables	19
Módulo 23	Danálisis de datos cualitativos	39
Módulo 24	Tabulación cruzada de los datos cuantitativos	63
Módulo 25	Medidas de la asociación con base en el riesgo	79
Módulo 26	Manejo de las variables de confusión	95
Módulo 27	Preparación del análisis estadístico	111
Módulo 28	Selección de una prueba de significación	127
Módulo 29	Determinación de las diferencias entre los grupos	
	Parte I: análisis de las observaciones sin aparear	145
Módulo 30	determinación de las diferencias entre los grupos, Parte II: Análisis de las observaciones apareadas	165
Módulo 31	Medición de las asociaciones entre las variables; regresión y correlación	175
Módulo 32	Redacción del informe de investigación	189
Módulo 33	Difusión, comunicación y uso de los hallazgos de la investigación	209
Acerca de los autores		223

Este espacio en blanco intencionalmente dejado de la página

PRÓLOGO

A lo largo de los últimos 20 años, la investigación sobre sistemas de salud (ISS) ha probado ser una herramienta útil para los responsables de la toma de decisiones en materia de salud a todos los niveles, debido a que provee los datos necesarios para la toma de decisiones sobre bases sólidas.

La Organización para el proyecto de la región sudafricana para la ISS –con base en la Oficina de la OMS en Harare– apoyada por la OMS en Ginebra, el Royal Tropical Institute (KIT) de Ámsterdam y el Dutch Technical Development Co-operation (DGIS), ha desempeñado un papel crucial en la promoción de la ISS en la región africana desde 1987. La ISS fue adoptada con entusiasmo por muchos de los ministerios de salud y diversas universidades. En 1996, la Oficina regional de la OMS para el África subsahariana (AFRO) asumió la responsabilidad de aplicar la ISS y, de acuerdo con las recomendaciones de los ministros de salud de la región, en 1998 la OMS/AFRO incluyó la ISS como un programa regular en los 46 estados miembros.

Los módulos de capacitación para la ISS que presentamos fueron elaborados por un equipo interdisciplinario internacional de investigadores empíricos, y han sido de gran utilidad para fomentar el interés en la ISS. A pesar de haber sido diseñados originalmente como una herramienta para los funcionarios del sector salud en el desarrollo de la investigación, orientada a la solución de problemas en la región sudafricana, los módulos también probaron su eficacia en Malasia y fueron ampliados por el personal de la Escuela de Salud Pública. La versión integrada de 1991, publicada por el Centro Internacional de Investigación para el Desarrollo (IDRC, por sus siglas en inglés), Canadá, y la OMS, Ginebra,¹ fue traducida al francés, al español y al portugués, y algunas secciones aparecieron también en árabe, en vietnamita y en chino. Alrededor del mundo, los módulos de capacitación para la ISS han facilitado la elaboración y la instrumentación de cientos de protocolos de investigación de trabajadores e investigadores de la

1 Corlien M. Varkevisser, Indra Pathmanathan y Ann Brownlee (1991) *Designing and conducting Health Systems Research projects. Part I: Proposal development and fieldwork; Part II: Data analysis and report writing*. Ottawa: División de Ciencias para la salud del Centro Internacional de Investigación para el Desarrollo (IDRC, por sus siglas en inglés) y Ginebra: Programa de Investigación y Desarrollo de Sistemas de Salud de la Organización Mundial para la Salud.

salud. Los módulos para la ISS son utilizados en las Comunidades de Salud y los departamentos de Ciencias Sociales de varias universidades africanas, asiáticas y latinoamericanas, a fin de capacitar a los estudiantes y prepararlos para el trabajo de campo. Han sido empleados también en cursos de Maestría en Salud Pública en Europa y EEUU, así como en programas internacionales de investigación dedicados a la investigación científica aplicada.

El uso imprevisto de estos módulos de capacitación en círculos tanto académicos como de la administración de la salud llevó a que la edición de 1991—y las reimpressiones subsecuentes— se agotaran rápidamente. Además algunos equipos de usuarios sugirieron cambios pertinentes; entonces la Unidad para la ISS en la AFRO, de común acuerdo con el IDRC, decidió mejorar los módulos de la ISS. Un grupo interdisciplinario de investigadores sudafricanos hizo la revisión y propuso algunas rectificaciones en dos talleres, uno en 1998 y el otro en 1999. Dos de los tres editores originales concluyeron la presente versión. El IDRC asumió la responsabilidad final de su publicación, que fue financiada por la AFRO y el propio IDRC, y editada por KIT.

Esperamos que esta última versión de los módulos tenga la misma trascendencia que las anteriores. Es una realidad que muchos de los problemas de salud —tanto los nuevos como aquellos que persisten—, necesitan con urgencia investigaciones funcionales. Es un reto importante sustentar y patrocinar las reformas en salud necesarias y, al mismo tiempo, asegurar la equidad en el acceso a los servicios de atención de la salud para los grupos de alto riesgo. La ISS es una de las herramientas con que contamos para comprender profundamente estos retos y enfocar de manera óptima nuestros recursos.

Dr. Rufaro R. Chatora, Director de: Division of Health Systems and Services Development (DSD), WHO/AFRO, Harare.

Dr. Christina Zarowsky, Especialista en Salud Senior, IDRC, Ottawa.

Ms. Catherine Hodgkin, Jefe del Departamento de Salud, KIT, Royal Tropical Institute, Amsterdam.

PREFACIO Y RECONOCIMIENTOS

El presente volumen, *‘Diseño y manejo de la investigación sobre sistemas de salud’* es una meticulosa revisión del Volumen 2 de las *Series de capacitación en sistemas de salud* publicadas en 1991 por el Centro Internacional de Investigación para el Desarrollo (IDRC, por sus siglas en inglés) en Canadá, y la sede de la OMS en Ginebra, editadas y reimprimas en varias ocasiones con el mismo nombre. Fue necesario revisar y mejorar sus módulos puesto que, con los años, se detectaron inevitables deficiencias y lagunas.

Algunos funcionarios de la salud, por ejemplo, indicaron que la aplicación de los hallazgos y las recomendaciones de la investigación estaba subexpuesta. Este punto fue considerado en el módulo 1, al que se le anexó una cuarta fase de aplicación de la investigación sobre sistemas de salud (ISS). Este ciclo de capacitación constaba de tres fases: elaboración de propuestas de ISS (15 días), trabajo de campo (alrededor de seis meses), y análisis de los datos y redacción del informe (dos semanas). Los temas de la aplicación de los hallazgos y la instrumentación de las recomendaciones de la investigación están más elaborados en el módulo 33. Los funcionarios de la salud solicitaron, además, abreviar los cursos, lo que es entendible. Nos ocupamos de este punto recalcando más explícitamente —en los módulos 1 y 3, así como en las Instrucciones del curso (anexas a la Parte 1 de este volumen), que la fase de elaboración de la propuesta de investigación puede acortarse si los equipos eligen el tema de investigación de campo antes de comenzar el curso, de preferencia con la guía de un facilitador. El Programa OMS/AFRO con sede en Harare ya está elaborando módulos para una rápida participación en la investigación en salud a nivel de los centros de salud y los distritos, participación que pueden desarrollarse e integrarse a las actividades diarias del personal y los miembros de la comunidad.

Los investigadores de Salud Comunitaria, Ciencias Sociales y otros departamentos universitarios e institutos de investigación en África subsahariana y otras partes del mundo, que están usando los módulos de capacitación, manifestaron otras inquietudes. Abogan por que, además de los ya muy subrayados enfoques de investigación analítica para la solución de los problemas, se dé un mayor peso a la investigación descriptiva. De aquí que en el módulo 4 se agregara una parte descriptiva al diagrama de análisis de los problemas. En todos los pasos subsiguientes de la investigación se hizo, siempre que fue preciso, la distinción entre los estudios analíticos y los descriptivos. También se dio un mayor peso a los

métodos de investigación cualitativa, que se integraron más con más profundidad a los métodos cuantitativos en la metodología de la investigación (módulos 8-14). Esto se hizo así, por ejemplo, en los módulos 10 (*Técnicas de recolección de datos*) y 11 (*Técnicas de muestreo*). Además, se agregaron dos módulos (*Análisis de los datos y Redacción del informe*) a la Parte 2 del volumen: uno sobre medidas de asociación con base en el riesgo (módulo 25), en el que se utilizaron partes del módulo 30 de la versión de 1991, y otro sobre el difícil tema de las variables de confusión (módulo 26). La necesidad de explayarse se reflejó también en la más reciente evaluación de la capacitación en la ISS (1997).¹

Finalmente, los facilitadores pidieron ejemplos más elaborados de las técnicas críticas de investigación y análisis de datos. De aquí que el módulo 10B (*Desarrollo de herramientas de investigación*) incluya una sección de técnicas para entrevistar, con ejercicios específicos, y el módulo 10C, un ejemplo de la transcripción de un grupo de discusión (GD) con anotaciones en el margen. A dicha solicitud se debe también que en el módulo 13 (*Plan para el análisis de los datos*), se haya agregado dos ejemplos, uno de un cuestionario completo y el otro de una hoja maestra. El módulo 23 (*Análisis de datos cualitativos*) incluye, asimismo, un ejemplo de una hoja de compilación completada. El módulo 5 (*Revisión bibliográfica*), fue ampliado con un ejemplo de revisión bibliográfica.

Además de estas adiciones, en todos los módulos se reescribieron las secciones que resultaron poco claras o incompletas, y muchos de los ejemplos y referencias se sustituyeron por otros más recientes o mejor elaborados.

Un grupo interdisciplinario de investigadores y funcionarios convocados por el director del Programa de ISS OMS/AFRO (desde 1992, Gabriel Mwaluko) comenzó la presente revisión en un taller realizado del 2 al 11 de noviembre de 1998 en Arusha. Todos los participantes tenían experiencia con los módulos y con la ISS: Sambe Duale, Lawrence Gakuri, Pilate Khulumani, Steve Kinoti, Gabriel Mwaluko, Jude Padayachi, Brian Pazvakavambwa, Corlien Varkevisser y Godfrey Woelk. En agosto de 1999, un grupo de tres personas (Alasford Ngwengwe, Corlien Varkevisser y Godfrey Woelk) hicieron las revisiones ulteriores y sincronizaron los diferentes textos en la oficina de la OMS/AFRO/ISS en Harare, apoyados por personal de la oficina de ISS (presidida desde 1999 por Isabel R. Aleta, con Makhamokha Mohale y Eric Naterop como APO). Corlien Varkevisser y Ann Brownlee terminaron los módulos y los editaron, con la aprobación de Indra Pathmanathan,

1 Corlien M. Varkevisser, Indra Pathmanathan y Ann Brownlee (1991) *Designing and conducting Health Systems Research projects. Part I: Proposal development and fieldwork; Part II: Data analysis and report writing*. Ottawa: División de Ciencias de la Salud del Centro Internacional de Investigación para el Desarrollo (IDRC, por sus siglas en inglés) y Ginebra: Programa de Investigación y Desarrollo de Sistemas de Salud de la Organización Mundial de la Salud.

quien en no pudo participar en esta ocasión. Deborah Karugonjo (Harare) y Merel Gallée (Ámsterdam) brindaron una muy valiosa ayuda en la producción de las sucesivas versiones computarizadas. La DGIS (Dutch Development Co-operation); SARA/AED, Washington; GTZ, Alemania, a través de la red GTZ MCH/FP para la Investigación sobre sistemas de salud en Sudáfrica; OMS/AFRO y la sede de la OMS en Ginebra proporcionaron los fondos para revisar y editar los módulos para la ISS. El IDRC, Canadá, apoyó la distribución subsidiada de los módulos.

Un grupo heterogéneo de personas apoyó la producción de las versiones anteriores de los módulos de la ISS. La cuna de los módulos ha permanecido en África Occidental, en donde el Proyecto para el Fortalecimiento de los Sistemas de Distribución de la Salud (SHDS, por sus siglas en inglés), con base en la Universidad de Boston, EEUU, produjo —a principios de la década de los ochenta, a solicitud de la AFRO— material de capacitación para la elaboración de protocolos de investigación. El SHDS siguió paso a paso el enfoque que a la fecha continúa siendo el más eficaz de los cursos de ISS. Los módulos 1-17 de este volumen son adaptaciones o nuevas versiones de los módulos originales del SHDS.² La primera adaptación tuvo lugar en 1988, con 12 investigadores de países que participaron en el Proyecto Común de ISS (Omondi (Kenya), Sebatane y Makatjane (Lesotho), Chimimba y Msukwa (Malawi), Kitua y Savy (Seychelles), Tembo (Zambia) Munochiveyi, Taylor y Woelk (Zimbabwe) y personal del Proyecto Común, que también terminó la versión (Corlien Varkevisser y Martien Borgdorff). Estos ‘módulos verdes’³ encontraron su camino a Malaysia, en donde Indra Pathmanathan continuó elaborándolos, con el apoyo de Maimunah Abdul Hamid, K. Mariappan y C.Sivagnanasundram (Sri Lanka), en el curso de numerosos talleres de elaboración de protocolos. Lo mismo sucedió en Sudáfrica y en África Occidental. A iniciativa de Yvo Nuyens, creador del Proyecto Común de ISS en la OMS, en Ginebra, con el apoyo del IDRC (Annette Stark), surgieron los cinco volúmenes de las *Serie de capacitación para la investigación sobre sistemas de salud*, de las que *Diseño y realización de proyectos de investigación en sistema de salud* conformó el Volumen 2. Estos “módulos rosas” publicados en Ottawa en 1991 por el IDRC y la OMS, constituyen una fusión completa de las versiones elaboradas en Sudáfrica y Malasia. Se integraron en Harare (Corlien Varkevisser y Leon

2 Regional Assessment of Health Systems Research Training in Eastern and Southern Africa, HSR Project and SARA/AED, Harare/Washington, SS Ndeki, 1997.

3 Ann Brownlee, Thomas Nchinda y Yolanda Mousseau-Gershman (1983). *Health Services Research Course: How to develop proposals and design research to solve priority problems*, Boston: Boston University Health Policy Institute.

Joint World Health Organisation/Royal Tropical Institute/Dutch Technical Development Co-operation Project on Health Systems Research for the Southern African Region (1988). *Health Systems Research Training Course: How to develop research proposals to solve priority health problems*, Geneva: World Health Organisation, WHO/SHS/HSR/88.3.

Bijlmakers), con la asesoría de Indra Pathmanathan y el meticuloso apoyo editorial de Ann Brownlee, una de las autoras de los módulos originales del SHDS. Los presentes módulos para la ISS son, por lo tanto, una producción mundial. Es imposible mencionar a todos los colaboradores, ya que varios facilitadores y participantes del curso de ISS también inspiraron modificaciones con sus preguntas y observaciones críticas.

Con antecedentes tan coloridos e interactivos, es poco probable que esta reimpresión sea la última. Mientras los módulos se utilicen, habrán de adaptarse. Esperamos, no obstante, que se mantengan vigentes algunos años en su forma actual y sean de utilidad tanto para los trabajadores de la salud como para los estudiantes universitarios.

Dr. Corlien M. Varkevisser, Royal Tropical Institute/University of Amsterdam

Dr. Ann Brownlee, University of California, San Diego

Junio de 2003

INTRODUCCIÓN A LA PARTE II

Análisis de los datos y redacción del informe

Esta publicación debe utilizarse en combinación con la Parte I: *Elaboración de la propuesta de investigación y trabajo de campo*. La Parte I consta de 20 módulos de capacitación que apoyan a los participantes del curso, paso a paso, en la elaboración de una propuesta de investigación y les brindan indicaciones útiles para su implementación.

La Parte II, *Análisis de los datos y redacción del informe*, que presentamos ahora, consta de 13 módulos de capacitación para analizar los datos, redactar el informe y planificar la implementación. Estos módulos pueden utilizarse de manera flexible, aún en mayor medida que los dedicados al desarrollo de la propuesta de investigación, dependiendo de:

- el nivel de escolaridad y la experiencia en el campo de la investigación de los participantes del curso
- el tipo de estudio realizado y el tipo de técnicas de recolección de datos empleadas
- el estado de los datos cuando se lleve a cabo el taller para su análisis y redacción del informe

Si los participantes ya tienen una cierta capacitación en métodos de investigación y en estadística, además de experiencia en la investigación, la presentación de los módulos puede ser breve. Su propósito será, entonces, refrescar la memoria de los participantes y guiarlos en la aplicación de los procedimientos apropiados de análisis y comprobación. Algunos módulos pueden combinarse o abreviarse.

Si los participantes no tienen ninguna capacitación ni experiencia en investigación, la presentación de los módulos podría tener que limitarse al material estrictamente esencial para manejar los datos recogidos. En este caso, las presentaciones pueden prolongarse y deben dar lugar a preguntas y ejercicios en clase.

“El material estrictamente esencial” incluye:

- El *módulo 21* (Orientación del taller)
- Los *módulos 22 y 24* (Descripción de las variables y Tabulación cruzada) completos
- El *módulo 23* (Análisis de los datos cualitativos, especialmente las *partes I a la IV*)
- El *módulo 25* (Medidas de asociación basadas en el riesgo: incidencia, riesgo, riesgo relativo y razón de probabilidades), concentrándose en las observaciones sin aparear
- *Brevemente: el módulo 26* (Manejo de las variables de confusión); empleando algunos ejemplos, debe quedar claro qué es la confusión y cómo manejarla
- El *módulo 27* (Preparación del análisis estadístico: medidas de dispersión, distribución normal y variación de la muestra)
- El *módulo 28* (Selección de una prueba de significación); concéntrese de lleno en las *partes I, II y III*, que explican el fundamento lógico de las pruebas de significación y cómo operan, pero trate sólo brevemente la parte IV, la selección propiamente dicha de una prueba de significación, si cree que los grupos sólo van a usar la prueba de chi cuadrado (χ^2) o la prueba t
- El *módulo 29* (Determinación de las diferencias entre los grupos: análisis de las observaciones sin aparear). Ya sea la prueba t, la prueba de chi cuadrado (χ^2), o ambas.
- Los *módulos 32* (Redacción del informe) y *33* (Promoción de la difusión, comunicación y uso de los hallazgos de la investigación) completos

Dependiendo del tipo de estudio que los participantes hayan realizado y del tipo de análisis necesario para sus datos, puede usted ampliar el alcance de las presentaciones (extenderse sobre las pruebas estadísticas, por ejemplo, o sobre el análisis de los datos cualitativos), o cambiar su secuencia (si los participantes tienen sobre todo datos cualitativos, puede presentar el *módulo 23* antes que el 22).

- Por lo general, la *primera mitad del taller* (una semana) se dedica a concluir el procesamiento de los datos y su análisis. Todos los módulos relacionados con el análisis (21-31) se presentan esta semana.
- El tiempo de estas presentaciones debe calcularse con todo cuidado. Los módulos 21-24 pueden presentarse antes de terminar de procesar los da-

tos, y los módulos 25 y 26 cuando los grupos estén por terminar el procesamiento de los datos y hayan terminado algunas de las tablas básicas. Continúe con el módulo 27, Preparación del análisis estadístico (medidas de dispersión, distribución normal y variación de la muestra).

- Los módulos que presentan diversas pruebas estadísticas deben exponerse cuando los participantes ya estén bastante adelantados en la preparación de las tabulaciones cruzadas.
- La *segunda mitad del taller* se dedica a la redacción del informe, la formulación de las recomendaciones, y la presentación y discusión en sesión plenaria de los principales hallazgos y recomendaciones derivados de los estudios. Por lo general esta semana hay sólo dos presentaciones: sobre la redacción del informe (módulo 32) y sobre la difusión, la comunicación y el uso de los hallazgos de la investigación (módulo 33). Es mejor presentar el último módulo justo antes de que los participantes formulen la síntesis de los hallazgos y las recomendaciones de sus estudios.

En las siguientes páginas se presenta, como ejemplo, el programa de un curso de dos semanas para el análisis de los datos y la redacción del informe.

Si los participantes tienen un buen nivel y procesaron satisfactoriamente los datos antes de que se convocara a este taller, el informe podría terminarse en un par de semanas, incluyendo la formulación de un plan de acción. De otro modo, los retoques finales deben hacerse después, con el apoyo de un facilitador, ya sea en persona o a distancia.

EJEMPLO DE PROGRAMA DE UN CURSO

(Como se utilizó en Sudáfrica)

Diseño y dirección de proyectos de ISS: análisis de los datos y redacción del informe

Día/hora	Sesión	Responsable(s)
<i>Lunes</i>		
08:00 – 08:30	Comentarios de apertura	Coordinador del curso
08:30 – 9:15	Presentación de los resultados preliminares y su discusión	Grupo 1
09:15 – 10:00	Presentación de los resultados preliminares y su discusión	Grupo 2
10:00 – 10:30	Té	
10:30 – 11:15	Presentación de los resultados preliminares y su discusión	Grupo 3
11:15 – 12:00	Presentación de los resultados preliminares y su discusión	Grupo 4
12:00 – 12:30	Módulo 21: Orientación para el taller de análisis de los datos y redacción del informe	Facilitador
12:30 – 14:00	Almuerzo	
14:00 – 15:30	Trabajo en equipo	
15:30 – 16:00	Té	
16:00 – 17:00	Módulo 22: Descripción de las variables	Facilitador
17:00 – 18:00	Trabajo en equipo	
<i>Martes</i>		
08:00 – 09:00	Módulo 23: Análisis de los datos cualitativos	
09:00 – 13:00	Trabajo en equipo (incluyendo el té)	Facilitador
13:00 – 14:00	Almuerzo	
14:00 – 15:00	Módulo 24: Tabulación cruzada de los datos cuantitativos	Facilitador
15:00 – 18:00	Trabajo en equipo (incluyendo el té)	
<i>Miércoles</i>		
08:00 – 13:00	Trabajo en equipo (incluyendo el té)	
13:00 – 14:00	Almuerzo	
14:00 – 15:00	Opcional: presentación de los principales resultados del trabajo en equipo: objetivos revisados, principales tablas cruzadas, resultados del análisis cualitativo	
15:00 – 16:00	Módulo 25: Medidas de asociación basadas en el riesgo (incidencia, riesgo, riesgo relativo, razón de probabilidades)	Los cuatro grupos
16:00 – 18:00	Trabajo en equipo (incluyendo el té)	Facilitador
<i>Jueves</i>		
08:00 – 09:00	Módulo 26: Manejo de las variables de confusión	Facilitador
09:00 – 13:00	Trabajo en equipo (incluyendo el té)	
13:00 – 14:00	Almuerzo	
14:00 – 15:00	Módulo 27: Preparación del análisis estadístico (medidas de dispersión, distribución normal y variación de la muestra)	Facilitador
15:00 – 18:00	Trabajo en equipo (incluyendo el té)	

Día/hora	Sesión	Responsable(s)
<i>Viernes</i>		
08:00 – 09:00	Módulo 28: Selección de una prueba de significación (20 minutos) seguido del módulo 29, partes I y II (prueba t)	Facilitador
09:00 – 13:00	Trabajo en equipo (incluyendo el té)	
13:00 – 14:00	Almuerzo	
14:00 – 15:00	Módulo 29, parte III (prueba de chi cuadrado (χ^2))	Facilitador
15:00 – 18:00	Trabajo en equipo (incluyendo el té)	
<i>Sábado</i>		
08:00 – 13:00	Trabajo en equipo (incluyendo el té) Módulos 30 o 31 (de ser necesario)	
<i>Domingo</i>		
Libre		
<i>Lunes</i>		
08:00 – 09:00	Módulo 32: Redacción del informe	Facilitador
El resto del día	Trabajo en equipo	
<i>Martes</i>		
Todo el día	Trabajo en equipo	
<i>Miércoles</i>		
08:00 – 08:30	Módulo 33: Promoción de la difusión, la comunicación y el uso de los hallazgos de la investigación	Facilitador
El resto del día	Trabajo en equipo	
<i>Jueves</i>		
08:00 – 13:00	Trabajo en equipo	
13:00 – 14:00	Almuerzo	
14:00 – 17:30+	Trabajo en equipo; preparación de la presentación de los resultados de la investigación, las recomendaciones y un plan de acción preliminar	Todos los grupos y los facilitadores
<i>Viernes</i>		
08:00 – 13:00	Trabajo en equipo para terminar los informes y las presentaciones	
13:00 – 14:00	Almuerzo con los invitados (funcionarios de la salud y la investigación interesados del Ministerio de Salud y la universidad e instituciones de investigación)	
14:00 – 17:00	Presentaciones de la síntesis de los hallazgos, las recomendaciones y el plan de acción tentativo, de los cuatro grupos, seguidas de la discusión (3/4 de hora por grupo, con descansos para el té, entre una y otra)	
17:00 – 17:30	<i>Evaluación</i> del Curso de Capacitación en ISS	
<i>Sábado (de ser necesario)</i>		
08:00 – 13:00	Retoques finales a los informes	

Este espacio en blanco intencionalmente dejado de la página

Módulo 21

ORIENTACIÓN PARA EL TALLER DE ANÁLISIS DE LOS DATOS Y REDACCIÓN DEL INFORME

Este espacio en blanco intencionalmente dejado de la página

Módulo 21 ORIENTACIÓN PARA EL TALLER DE ANÁLISIS DE LOS DATOS Y REDACCIÓN DEL INFORME

OBJETIVOS DEL TALLER

Al terminar este taller, usted debe ser capaz de:

1. *Identificar* y definir los conceptos básicos y los procedimientos necesarios para el análisis y la interpretación de datos.
2. *Analizar* e interpretar los datos recogidos para el proyecto de investigación que desarrolló en el primer taller, y formular conclusiones con respecto a los objetivos de su estudio.
3. *Redactar* un informe de investigación claro y conciso, y una síntesis de los principales hallazgos y recomendaciones para cada uno de los grupos interesados en los resultados.
4. *Presentar* los principales hallazgos y las recomendaciones de su estudio a los responsables de la toma de decisiones y a las personas objeto de la investigación, para terminar de hacer las recomendaciones.
5. *Preparar* un plan de acción para difundir, comunicar y utilizar los hallazgos y, dado el caso, proponer otras investigaciones.

I. Revisión de la experiencia de campo.

II. Introducción al taller.

III. Tareas que deben llevarse a cabo en el taller.

1. Revisión y conclusión del procesamiento de datos.
2. Análisis de los datos.
3. Redacción del informe.
4. Presentación de una síntesis de los hallazgos y las recomendaciones.
5. Formulación de un plan para implementar los resultados de la investigación.

I. REVISIÓN DE LA EXPERIENCIA DE CAMPO

Instrumentar la propuesta de investigación debe haber representado un gran reto para usted. Sin duda debió enfrentar más de un obstáculo imprevisto conforme se adentró en el trabajo de campo, pero también habrá disfrutado el éxito. Si todo ocurrió según el plan, ya recogió sus datos, procesó gran parte de ellos, si no es que todos, comenzó su análisis y escribió un informe preliminar de las experiencias y los resultados de su trabajo de campo. Sus experiencias prácticas en la conducción del proyecto son invaluable. Compartir las con los demás en este taller va a ser un ejercicio muy provechoso, puesto que tanto sus problemas como sus éxitos pueden significar valiosas lecciones para el futuro.

Antes de dar un panorama general del enfoque y las actividades de este taller, reserve un tiempo para oír las experiencias de cada uno de los grupos de investigación.

EJERCICIO: Presentaciones de las experiencias de campo

Presente el informe preliminar que su grupo preparó al final de la experiencia de campo, de acuerdo con las indicaciones expuestas en el *módulo 20*. Prepárese para responder a las preguntas de los demás participantes y los facilitadores al final de la presentación.

Cada grupo va a disponer de 10 a 15 minutos para su presentación y de unos minutos más para preguntas y discusión.

II. INTRODUCCIÓN A ESTE TALLER

Este taller es continuación de aquel en que elaboró usted su propuesta de investigación. Ahora tiene por delante la importante tarea de analizar los datos recogidos en el trabajo de campo y redactar su informe de investigación. El informe debe tener recomendaciones factibles y útiles, derivadas de los hallazgos de su estudio, para resolver el problema investigado.

Al igual que en el primer taller, va a haber presentaciones y sesiones de trabajo en equipo y plenarias. Pero ahora el trabajo en equipo va a ocupar la mayor parte del tiempo. Las presentaciones se van a concentrar en la primera semana, que se va a dedicar al análisis de los datos. La segunda semana se va a reservar por completo para la redacción del informe, con sólo dos presentaciones para orientarla. Para finales de esa semana, cada grupo va a presentar una síntesis de sus principales hallazgos y recomendaciones en una sesión plenaria. A esta sesión se va a invitar a un selecto grupo entre los responsables de la toma de decisiones y los funcionarios de la salud que solicitaron el estudio o que estén directamente

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none"> – Revise la experiencia de campo – Haga un inventario de los datos de cada uno de los objetivos/población de estudio – Clasifique los datos y revise su calidad – Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none"> – Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none"> – Tablas cruzadas analíticas (24) – Medidas de la asociación basada en el riesgo (25) – Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none"> – Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none"> – Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none"> – Pruebas t y chi cuadrado (χ^2) (29) ** prueba t para datos apareados **prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none"> ** Diagrama ** Línea de regresión ** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none"> – Prepare un bosquejo del informe – Presente e interprete los datos – Formule y reformule – Comente y sintetice las conclusiones – Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none"> – Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

interesados en el tema, así como a investigadores que lo soliciten, para comentar su presentación.

Los *módulos* de este taller abarcan varias tareas importantes, que se presentan esquemáticamente en el diagrama de la siguiente página. Este diagrama se vuelve a presentar al principio de cada uno de los módulos subsecuentes para indicar la tarea que enfoca la presentación. Ahora vamos a dar una breve introducción a estas tareas.

I. TAREAS QUE DEBEN LLEVARSE A CABO EN EL TALLER

1. Revisión y conclusión del procesamiento de datos

Aunque confiamos en que todos los grupos han hecho un gran esfuerzo para procesar sus datos, puede ser necesario hacer algún ajuste. Esto es normal en la investigación: cada paso hacia adelante puede ir seguido de medio paso atrás. Incluso en una etapa bastante adelantada del análisis de los datos, puede usted tener que reagrupar y volver a procesar algunos de ellos.

Antes de empezar el análisis de los datos, se debe revisar si su procesamiento se hizo de manera tal que la información:

- sea fácil de manejar
- no tenga ya ningún error cometido en la recolección

Debe usted hacerse, por lo tanto, las siguientes preguntas:

- *¿Todos los datos se clasificaron adecuadamente? ¿Los cuestionarios y las listas de verificación se numeraron de la manera más conveniente? ¿Las principales categorías de informantes (por ejemplo, h/m, casos y controles) se distinguen en forma clara de modo de facilitar la comparación en las variables relevantes, tal como lo requieren los objetivos de la investigación?*
- *¿Se hizo el control de calidad de todos los datos para asegurar que la información esté completa y sea consistente? Vea en el *módulo 13* las medidas que debe tomar si los datos están incompletos o son inconsistentes.*
- *¿Todos los datos se ingresaron en la computadora o se transcribieron en hojas maestras? ¿El total de respuestas corresponde al total de personas encuestadas para cada una de las variables? De no ser así, ¿se pasaron por alto datos perdidos?*
- *¿Los datos cualitativos se categorizaron hasta donde fue posible? ¿Se terminó la codificación, en su caso? (Consulte el *módulo 13* sobre la post-ca-*

tegorización de las preguntas abiertas). ¿Los GD se leyeron y se ordenaron cuidadosamente de acuerdo con los temas de discusión? ¿Se destacaron las partes o las preguntas más ilustrativas de los objetivos de la investigación? Puesto que en principio la recolección, la clasificación, la síntesis y el análisis de los datos cualitativos se entrecruzan (vea el *módulo 10C*), usted ya debe haber interpretado gran parte de sus datos cualitativos. El *módulo 23* va a retomar el análisis pormenorizado de los datos cualitativos.

- Si usted utilizó la computadora para procesar sus datos, *revise la cuenta de la frecuencia de cada una de las variables* del cuestionario. Revise también las tabulaciones cruzadas. En el *Anexo 21.1* se describe pormenorizadamente cómo hacerlo.

Antes de revisar los procedimientos para procesar los datos, haga usted un INVENTARIO de todos los datos disponibles para cada OBJETIVO. Esto es particularmente importante cuando se emplean distintas herramientas en la recolección de datos.

Por ejemplo:

Las fuentes de datos para el objetivo 3: ‘Detección de las deficiencias en el funcionamiento de los servicios materno infantiles que expliquen la escasa utilización de los servicios de atención del parto:’

- Cuestionario para las madres, preguntas 12, 15 – 19, 23
- GD con trabajadores de la salud, temas 3 y 4
- Observaciones incluidas en las listas de verificación

Un inventario semejante lo ayuda a organizar mejor el análisis de los datos y, más adelante, a redactar el informe.

2. Análisis de los datos

Al comenzar el análisis de nuestros datos, debemos considerar cuáles son cuantitativos y cuáles cualitativos.

Datos cuantitativos

Los datos cuantitativos se expresan en números y se presentan, por lo general, en tablas de frecuencia. De sus hojas maestras de datos puede usted fácilmente derivar los totales para cada variable/pregunta, contar la cantidad de respuestas obtenidas y presentar la información en tablas de frecuencia. (Véase el *módulo*

22.) Al analizar los datos cuantitativos, es importante considerar el propósito del estudio. Éste es:

- *¿Describir determinadas variables?*
Por ejemplo: la distribución de las adolescentes embarazadas en una determinada población
- *¿Buscar diferencias entre grupos?*
Por ejemplo: diferencias en el ingreso o en el estado de salud entre los antiguos pobladores y los recién llegados en un área determinada
- *¿Determinar una cierta asociación entre las variables?*
Por ejemplo: la asociación entre la satisfacción de las enfermeras en el trabajo y la cantidad de reuniones para los empleados realizadas el año anterior

Las tabulaciones cruzadas son una herramienta importante para sintetizar y analizar este tipo de datos (*módulo 24*), aunque hay otras opciones (vea los *módulos 22, 25 y 31*).

Una vez que haya hecho las distribuciones de frecuencias y diferentes tipos de tabulaciones cruzadas, debe elegir el tipo de análisis estadístico necesario para determinar si las diferencias y las asociaciones encontradas son significativas o mero producto de la casualidad. La selección de las pruebas de significación apropiadas se explica pormenorizadamente en el *módulo 28*. En los *módulos 29-31* se presentan algunos conceptos estadísticos más avanzados para el análisis de los datos cuantitativos.

Las pruebas de significación más comunes son:

- La prueba *t* de Student y la prueba de chi cuadrado (χ^2), para determinar las diferencias entre diversos grupos cuando las observaciones no están apareadas (*módulo 29*).

La prueba *t* para datos apareados y la prueba χ^2 (chi cuadrado) de McNemar, para determinar las diferencias entre grupos cuando las observaciones están apareadas (*módulo 30*).

Para medir las asociaciones entre las variables, se presentan los conceptos de razón de probabilidades (*módulo 26*) y de regresión y correlación (*módulo 31*).

A lo largo del proceso de análisis de los datos, es importante tener en mente que *nuestros hallazgos deben responder a las preguntas de investigación para así alcanzar sus objetivos. Vamos, finalmente, a esbozar conclusiones y a recomendar acciones basadas en estos hallazgos.*

Datos cualitativos

Recuerde que puede obtener datos cualitativos mediante:

- *preguntas abiertas* no pre-categorizadas, en cuestionarios o programas de entrevista que también recojan datos cuantificables
- *entrevistas de estructura flexible* y predominio de preguntas abiertas dirigidas a informantes clave (individuos) o a grupos reducidos
- *grupos de discusión* sobre temas seleccionados, con listas de puntos a tratar para guiar la discusión
- *observaciones* que describan el comportamiento individual o del grupo
- *diarios, ensayos* y cualquier información que se haya generado con *métodos proyectivos* (por ejemplo, oraciones incompletas, relatos con una laguna, asociaciones libres de los informantes al mostrárseles fotografías o películas)

Como recordará del ejercicio que hizo en el *módulo 13*, las respuestas a las preguntas abiertas se pueden:

- *enlistar*
- *categorizar* (de acuerdo a los objetivos de investigación y al sentido común, combinando las respuestas que “encajan” juntas entre 4 y 6 categorías, rara vez más)
- *codificar/ etiquetar*
- *interpretar* por categoría de contenido, dependiendo del propósito para el que necesita los datos
- *insertar*, utilizando estos códigos, *en sus hojas maestras o en la computadora*
- *contar*, como otros datos cuantitativos

Puede usted haber procesado ya algunas preguntas abiertas bien definidas. Vamos a profundizar en estos procedimientos en el *módulo 23*, para resolver las dudas al respecto.

Nota:

La principal característica del análisis de los datos cualitativos es que éste emplea PALABRAS y no números.

Los datos cualitativos de otras fuentes –distintas de las preguntas abiertas– requieren técnicas de codificación y compilación más complejas. Con frecuencia es útil sintetizar los datos cualitativos en hojas de compilación, diagramas, diagramas de flujo o matrices que nos ayudan en el análisis. El módulo 23 trata el análisis de dichos datos cualitativos.

3. Redacción del informe

Usted volverá a casa con un informe completo de su investigación, con los siguientes componentes:

1. Una INTRODUCCIÓN que cubra el planteamiento del problema, datos contextuales relevantes y una revisión bibliográfica.
2. OBJETIVOS
3. Una sección sobre el MÉTODO, que informe de cuándo, dónde, y cómo recogió los datos, cómo los analizó y las posibles flaquezas de la recolección y el análisis.
4. HALLAZGOS
5. DISCUSIÓN
6. CONCLUSIONES y RECOMENDACIONES

Las tres últimas secciones, que van a formar el grueso del informe, se comentan en el *módulo 32*. Puede usted revisar y sintetizar las tres primeras secciones a partir de las secciones correspondientes de su propuesta de investigación.

4. Presentación de una síntesis de los hallazgos y las recomendaciones

Puesto que una meta importante de su investigación es que se emprendan acciones pertinentes basadas en los resultados de su estudio, es importante que todos los sectores interesados tengan oportunidad de comentar los hallazgos y las recomendaciones antes de que concluya usted el informe. Puede usted querer incluir a quienes formulan las políticas, a los funcionarios de la salud, a miembros del personal y de la comunidad, e incluso a los medios de comunicación. El *módulo 33* presenta indicaciones para organizar reuniones con este propósito.

5. Formulación de un plan para instrumentar los resultados de la investigación

En el taller anterior (*módulo 17*) formuló un plan para utilizar y difundir sus resultados. Ahora va a revisar y a desarrollar meticulosamente este plan, incluyendo la planificación de la implementación de las recomendaciones derivadas de su estudio (*módulo 33*).

Trabajo en equipo (Tiempo flexible, dependiendo de los temas de investigación y el estado del procesamiento de los datos)

- Reconsidere los objetivos de su propuesta de investigación y *liste las diferentes fuentes de datos para cada objetivo* (cuestionarios, registros, grupos de discusión, etc.).
NB: Si descubre que para explicar su problema de investigación recogió más datos de los necesarios para sus objetivos, puede usted revisarlos y agregar uno o dos más. Pero si, en cambio, recogió menos datos de los necesarios, no descarte los objetivos que no pueda abordar; explique, más bien, por qué no pudo hacerlo.
- Verifique que todos los datos estén completos, que sean consistentes y que estén correctamente codificados. Si no lo están, corrijalos.
- Determine si las hojas maestras se prepararon para diferentes poblaciones de estudio o para diferentes categorías de informantes que quisiera comparar. Esto va a facilitar el análisis.
- Puede también marcar los cuestionarios de sub-grupos con diferentes colores de manera que pueda volver a los datos en crudo al revisar una determinada pregunta.
- Revise si las hojas maestras están completas y si el número de respuestas de cada variable concuerda con el número de personas encuestadas.
- Determine si se transcribieron y se ingresaron en la computadora todos los datos (véase el *anexo 21.1*).
- Revise si los datos cualitativos se categorizaron y sintetizaron en el campo. Si no, léalos y ordénelos, y ponga en los márgenes números por tema de discusión y palabras clave adicionales. En el módulo 23 se encuentran más indicaciones.
- Conforme enliste sus diferentes fuentes de datos por objetivo, revise si registró todas las observaciones pertinentes.

ANEXO 21.1: Salida de la computadora

La copia impresa en la computadora es el resultado de los comandos empleados en los programas de cómputo para analizar los datos disponibles. La precisión de la información depende, por lo tanto, de:

- los datos ingresados
- los programas empleados

El dicho “si entra basura, sale basura” es muy adecuado para los procesos que se realizan en la computadora. La precisión de los datos es responsabilidad tanto del equipo de investigación como del especialista en informática.

Tipos de copias impresas en la computadora

1. Lista de datos

La copia impresa de la lista de los datos ingresados en la computadora es útil cuando para hacer correcciones en los datos en el proceso de validación.

2. Distribución de frecuencias

Esta copia nos proporciona una distribución de frecuencias de las variables del cuestionario (con porcentajes). Vea el siguiente ejemplo de una distribución de frecuencias y considere cómo se relaciona con los cuestionarios.

Para asegurar que se utilicen los programas correctos, el especialista en computación debe estar familiarizado con el formato del cuestionario empleado, y con el proceso de recolección de datos.

Ejemplo:

¿La estancia hospitalaria fue adecuada? (n = 3306)

Sí	No	No sabe	Total
1937	1369	0	3306

¿La admisión fue adecuada? (n = 3306)

Sí	No	No sabe	Total
634	719	16	1369

En este ejemplo, el total de las respuestas: “sí”, “no” y “no sabe” a la pregunta “¿La estancia hospitalaria fue adecuada?” es igual al total de los casos estudiados. El total de respuestas a la pregunta “¿La admisión fue adecuada?” es de sólo 1369. El especialista en computación debe estar consciente de que esta pregunta se les hizo únicamente a los informantes que se mostraron insatisfechos con su estancia en el hospital, de otro modo la impresión de la computadora va a indicar que faltan 1937 respuestas, y no que éstas son las que no corresponden o aplican (NA).

Deben obtenerse distribuciones de frecuencias para cada una de las preguntas del cuestionario. Utilice la distribución de frecuencias para asegurar que:

- el total de personas encuestadas en cada pregunta sea el correcto (es decir, debe coincidir con las personas de la muestra a quienes se les preguntó)
- todos los códigos son relevantes para la pregunta. *Por ejemplo*, no debe haber códigos 3-8 en una pregunta que sólo tiene dos respuestas posibles (por ejemplo, sexo: M o F) y un código para “se ignora” (a “se ignora” se le da por lo general el código 9).

3. Tabulación cruzada

La copia impresa de la *tabulación cruzada* muestra el número de individuos que tienen dos (o más) de las variables estudiadas.

Ejemplo:

	Hombre	Mujer	Total
Enfermo			
No enfermo			
Total			

Antes de usar la tabulación cruzada, revise lo siguiente:

- El gran total de la tabla debe corresponder al número de individuos de la muestra
- Los totales de las columnas y las filas deben corresponder a la distribución de frecuencias de cada variable (es decir, el número de hombres y de mujeres debe corresponder a las respectivas distribuciones de frecuencias)
- De igual modo, el número de “enfermos” y “no enfermos” debe corresponder a la distribución de frecuencias. Si no es así, lo más probable es que haya un error en el programa. Consulte a su especialista en computación.
- Si en la copia impresa de la computadora hay una indicación de “casos perdidos,” eso significa que:
 - hay un código equivocado en la captura de los datos (por ejemplo, un código 4 cuando sólo puede haber 1, 2 o 9), o
 - las categorías que especificó no son exhaustivas.

Por ejemplo:

El cuestionario incluyó la opción “se ignora,” pero el programa de cómputo no. Por consiguiente, todos los casos de “se ignora” aparecen como “casos perdidos.”

El estado civil en el cuestionario incluyó: “casado”, “soltero”, “divorciado” y “viudo”; sin embargo, el programa de cómputo tiene registrados solamente: “casado”, “soltero” y “divorciado”, de modo que todas las personas viudas aparecen como casos perdidos.

Si las categorías por edad son de 10 a 14 y de 15 a 19 pero el programador las categoriza accidentalmente como de 10 a 13 y de 15 a 19, van a faltar todos los individuos de 14 años.

Notas del instructor

Módulo 21 ORIENTACIÓN PARA EL TALLER DE ANÁLISIS DE LOS DATOS Y REDACCIÓN DEL INFORME

Tiempo estimado y recursos didácticos

3 horas	Presentación de las experiencias de campo
¾ hora	Introducción y discusión
1 hora+	Trabajo en equipo (duración dependiendo de los temas de investigación y el estado del procesamiento de los datos)

Introducción y discusión

- Dedique la primera parte de la sesión introductoria al informe de los participantes de su experiencia de campo. Si todos los grupos llegaron preparados para presentar su informe preliminar, la sesión puede comenzar directamente con esta actividad. Si los grupos todavía necesitan tiempo para preparar su presentación, debe dársele ya sea antes o al principio de esta primera sesión.
- En la introducción al taller debe subrayar que si bien hay varias tareas por terminar –de entre las cuales el análisis de los datos y el informe de los hallazgos son las que más tiempo van a consumir–, el propósito último de los proyectos de investigación sigue siendo la preparación de las recomendaciones y su implementación. Puede pedirles a los participantes *sugerencias* con respecto a qué *personas responsables de formular políticas y qué directores se debe invitar* a la presentación y discusión de sus hallazgos y sus recomendaciones al final del taller.
- Al presentar el diagrama, considere utilizar diferentes láminas transparentes sobrepuestas.

- Adecue la presentación al nivel y los intereses de los participantes. Re-frésqueles la memoria con ejemplos del procesamiento de las preguntas abiertas y explíqueles la diferencia entre los estudios descriptivos, la comparación de grupos, los estudios que indagan diferencias entre dos o más grupos, y los estudios que determinan las asociaciones entre dos o más variables *—de preferencia con ejemplos de su propia investigación.*
- No asuste a los grupos que tengan poca experiencia en estadística con pormenores de las pruebas; sólo mencione que cada tipo de estudio necesita diferentes pruebas.
- Subraye la importancia de enlistar todos los datos disponibles para cada objetivo, incluyendo los datos cualitativos. Conforme avance el taller, se va a poner tal énfasis en la preparación de las tablas, que los participantes van a tender a olvidar observaciones valiosas e información obtenida con los informantes clave. El facilitador debe pedirles que registren esta información (si es que no lo han hecho ya) y la incluyan en la lista de datos disponibles para cada objetivo. Cuando estén escribiendo el informe, revise que ésta haya sido analizada.

Trabajo en equipo

- Lea las indicaciones para el trabajo en equipo con los miembros del grupo. Permítales re-examinar sus objetivos, liste los datos disponibles para cada objetivo y comente si los objetivos son lo bastante específicos para cubrir todos los datos relevantes recogidos. A veces, para facilitar el análisis, los objetivos deben separarse, re-frasearse, agregarse, o reordenarse. *Nunca* permita que el grupo omita un objetivo sin una explicación (en la sección dedicada al método) de por qué no puede considerarse.
- Revise con los miembros del grupo que todos los datos disponibles estén completos, exentos de errores, etc. Cerciórese de que hayan preparado hojas maestras por separado para las diferentes poblaciones del estudio o para los diferentes sub-grupos que se van a comparar, o que los datos de los diferentes sub-grupos puedan recuperarse fácilmente de la computadora.

Dese un tiempo extra, como facilitador, para introducir todos los datos disponibles, detectar posible flaquezas, y considerar diversas posibilidades para el análisis. De no hacer esto al comienzo del taller, va a ser muy difícil que pueda guiar eficientemente a los grupos para que logren resultados óptimos con los datos que recogieron.

- Los miembros del grupo pueden trabajar en sub-grupos para terminar de procesar los datos, pero cerciórese de comentar cada tanto los problemas y los avances con todo el grupo.

Este espacio en blanco intencionalmente dejado de la página

Módulo 22

DESCRIPCIÓN DE LAS VARIABLES

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 22 DESCRIPCIÓN DE LAS VARIABLES

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Describir los datos* en términos de distribuciones de frecuencias, porcentajes y proporciones.
2. *Usar* gráficos para presentar datos.
3. *Explicar* la diferencia entre media, mediana y moda.
4. *Calcular* la frecuencia, los porcentajes, la proporción, la razón, la tasa, la media, la mediana y la moda de las principales variables de su estudio que requieran dichos cálculos.
5. *Identificar* otras variables independientes (además de las identificadas en el primer taller), si las hubiera, necesarias para el análisis de sus datos.

I. Introducción.

II. Distribuciones de frecuencias.

III. Porcentajes, proporciones, razones y tasas.

IV. Gráficos.

V. Medidas de tendencia central.

I. INTRODUCCIÓN

Cuando seleccionó las variables de su estudio, en el *módulo 8*, lo hizo desde el supuesto de que esto lo ayudaría a definir su problema (variables dependientes) y sus distintos componentes, o los factores que contribuían a fomentarlo (variables independientes). El propósito del análisis de los datos es comprobar o descartar dichos supuestos, y presentar nuevas perspectivas del problema estudiado. *El propósito último del análisis es responder con sus datos a las preguntas de la investigación esbozadas en los objetivos.*

Pero antes de ver cómo pueden estar afectándose las variables entre sí, es necesario sintetizar la información obtenida para cada variable en una forma tabular simple o en una ilustración.

Algunas de las variables pueden haber producido datos numéricos, y otras, datos categóricos. Al analizar nuestros datos debemos determinar, antes que nada, el tipo de datos que estamos manejando. Esto es crucial porque el tipo de datos más usados determina el tipo de técnicas estadísticas que deben emplearse para probar³. si los resultados del estudio son significativos.

Los datos categóricos

Hay dos tipos de datos categorizados: nominales y ordinales (véase el *módulo 8*).

Cuando lo datos son *nominales*, las variables se dividen en un determinado número de categorías que no pueden ordenarse unas por encima de otras (puesto que no son unas mayores y otras menores).

Por ejemplo:

Datos nominales	Categorías
Sexo	hombre, mujer
Estado civil	soltero, casado, viudo, separado/divorciado

Cuando los datos son *ordinales*, las variables se dividen en un determinado número de categorías que se ordenan de menor a mayor o viceversa.

Por ejemplo:

Datos ordinales	Categorías
Nivel de conocimiento Opinión de una declaración	Bueno, regular, malo Completamente de acuerdo, de acuerdo, duda, en desacuerdo, totalmente en desacuerdo

Los datos numéricos

Hablamos de datos *numéricos* cuando éstos se expresan en números

Los hay dos tipos: discretos y continuos.

Los datos *discretos* son una serie de números definida.

Por ejemplo:

Datos discretos	Valores
Número de accidentes en vehículos motorizados	0, 1, 2, 6, 19, etc.
Número de visitas a la clínica	2, 4, 10, 0, 3, etc.
Número de embarazos por mujer	2, 12, 5, 0, 5, 4, etc.

Los datos *continuos* vienen de variables que pueden medirse con mayor exactitud, dependiendo de la precisión de los instrumentos de medición, y cada valor puede crecer o decrecer ilimitadamente.

Por ejemplo:

Datos continuos	Valores
Estatura (dos decimales)	12.12, 9.95, 45.13, 6.99, 28.78, etc.
Temperatura (en grados Celsius)	35.5, 37.8, 39.2, 40.1, 36.9, etc.
Edad (cumplida)	50, 45, 12, 78, 25, 16, 61, 90, etc.

Los datos numéricos pueden presentarse como:

- Distribuciones de frecuencias
- Porcentajes, proporciones, razones y tasas

- Gráficos
- Medidas de tendencia central

Vamos a comentar ahora estas operaciones una por una tanto para los datos categóricos como para los numéricos.

II. DISTRIBUCIONES DE FRECUENCIAS

Una *distribución de frecuencia* es la descripción de los datos presentada en forma tabular de modo que los datos sean más manejables. Nos da la frecuencia con que un determinado valor aparece en los datos.

Cuando en su proyecto de investigación, usted contó en sus hojas maestras el número de respuestas de cada categoría, hizo la distribución de frecuencias para todas sus variables. Ahora vamos a exponer brevemente algunos puntos importantes.

1. Los datos categóricos pueden tener categorías muy simples.

Ejemplo 1:

Para identificar los métodos de planificación familiar empleados por las adolescentes del distrito de Kweneng, en Botsuana occidental, se les preguntó cuál estaban empleando.

Los resultados se presentan en la siguiente *distribución de frecuencia*:

Método	Número
Abstinencia	14
Condomes	47
Inyectables	1
Implante anticonceptivo	1
Pastilla	35
Ninguno	307
Total	405

Estos datos son *nominales*. La distribución de frecuencia se calcula con tan solo obtener el total de respuestas en cada categoría.

Se debe revisar siempre que el total de respuestas concuerde o coincida con el número de personas encuestadas. De ser necesario, debe crearse una categoría de respuestas faltantes.

Las distribuciones de frecuencias se expresan, por lo general, en porcentajes (véase la parte III de este *módulo*). Viendo la distribución de frecuencia del ejem-

plo anterior, podemos concluir que alrededor de 75% de las adolescentes, o tres de cada cuatro, no utilizan métodos de planificación familiar. Entre aquellas que sí los usan, los más comunes son los condones y las pastillas.

Ejemplo 2:

Se preguntó al personal de salud de 148 instituciones rurales de salud: ¿Con qué frecuencia se ha quedado usted sin medicamentos para el tratamiento de la malaria en los últimos dos años? Ésta fue una pregunta cerrada con las siguientes opciones de respuesta: nunca, 1 o 2 veces (rara vez), 3 a 5 veces (ocasionalmente), más de 5 veces (frecuentemente).

El total de respuestas en cada categoría arroja la siguiente distribución de frecuencias:

Categoría	Número
Nunca	47
Rara vez	71
Ocasionalmente	24
Frecuentemente	6
Total	148

En este ejemplo, los datos son *ordinales*. El orden de las categorías es importante ya que éstas indican, de arriba abajo, la creciente gravedad del problema.

Los resultados de la distribución de frecuencia indican que la mayoría de las clínicas rara vez se quedan cortas de medicamentos contra la malaria, pero que éste es un problema ocasional en casi una sexta parte de las clínicas, y grave en unas cuantas.

2. Datos numéricos

Los procedimientos para hacer distribuciones de frecuencias de datos numéricos son muy similares a los empleados para los datos categóricos, excepto que ahora los datos deben agruparse en categorías. Los pasos que deben seguirse para hacer una distribución de frecuencia son los siguientes:

1. Seleccione grupos para clasificar los datos
2. Cuente el número de medidas en cada grupo
3. Súmelas y revise los resultados

La manera en que se seleccionan los grupos puede afectar cómo se van a ver los resultados. No nos podemos valer más que del sentido común, pero si cabe la sospecha de que estemos ocultando información por una mala selección de grupos, tendremos que cambiar las agrupaciones.

Ejemplo 3:

En los centros de salud del distrito X se está presentando un gran número de casos de malaria y usted desea sintetizarlos. Compare los resúmenes diarios y los semanales de los mismos datos, tal como se muestra en el *cuadro 22.1*:

Tanto los datos diarios como los semanales muestran un número creciente de casos de malaria, pero la situación de mejora mostrada los días 19, 20 y 21 no se refleja en el resumen semanal. Por lo tanto, sería mejor utilizar los datos diarios para indicar cuándo exactamente comenzaron a descender los números de los casos de malaria reportados.

Cuadro 22.1 Resúmenes diarios y semanales de los casos de malaria reportados en los centros de salud del distrito X

Día 1	9 casos	Semana 1	88 casos
Día 2	12		
Día 3	11		
Día 4	13		
Día 5	14		
Día 6	13		
Día 7	16		
Día 8	16 casos	Semana 2	131 casos
Día 9	16		
Día 10	18		
Día 11	19		
Día 12	16		
Día 13	21		
Día 14	25		
Día 15	28	Semana 3	168 casos
Día 16	28		
Día 17	28		
Día 18	32		
Día 19	21		
Día 20	19		
Día 21	12		

Reglas para la agrupación de datos:

- Los grupos no deben superponerse; de otro modo habrá confusión respecto de a qué grupo corresponde una determinada medida.

- Debe haber continuidad de un grupo al siguiente; es decir que no debe haber brechas, de otro modo algunas medidas pueden no encajar en ningún grupo.
- Los grupos deben ir de las medidas más bajas a las más altas, y todas deben tener un grupo al que puedan asignarse.
- Los grupos deben ser, por lo general, de la misma amplitud, de modo que los números de los diferentes grupos puedan compararse fácilmente

Sin embargo, a veces es válido elegir grupos de diferente amplitud como, por ejemplo, cuando nos interesamos en grupos de edad específicos (por ejemplo, menores de 1 año, de 1 a 4 años, de 5 a 14 años).

Cuando comience a sintetizar los datos, considere que es mejor formar grupos de más y no que resulten escasos. Así, durante el análisis de los datos puede usted combinar los grupos para formar nuevas categorías sin tener que revisar todos sus datos. Si, en cambio, tiene muy pocos grupos y desea formar más grupos, debe regresar a sus datos crudos.

Un mayor número de grupos da, por lo general, un cuadro más preciso, aunque si se forman demasiados grupos se puede perder la perspectiva general.

Como regla general elija números redondos para los valores en el límite inferior del grupo.

Por ejemplo: 1.00–9.99, 10.00–19.99, 20.00–29.99, o:
0–4; 5–9, 10–14, etc.

III. PORCENTAJES, PROPORCIONES, RAZONES Y TASAS

1. Porcentajes

En vez de presentar los datos en tablas de frecuencia utilizando números absolutos, muchas veces es preferible calcular porcentajes.

Un PORCENTAJE es el número de unidades con una determinada característica en la muestra, dividido entre el total de unidades de la muestra y multiplicado por 100.

Los porcentajes también se llaman FRECUENCIAS RELATIVAS. Estandarizan los datos, lo que facilita compararlos con datos similares obtenidos en alguna otra muestra de diferente tamaño u origen.

Ejemplo 4:

En 82 clínicas de un distrito se pidió al personal que presentara el número de pacientes tratados de malaria en un mes. Los investigadores presentaron tanto la distribución de frecuencia como los porcentajes (o frecuencias relativas):

Cuadro 22.2 Distribución de las clínicas de acuerdo con el número de pacientes tratados de malaria en un mes

Número de pacientes	Número de clínicas	Frecuencia relativa %
0 a 19	25	31
20 a 39	3	4
40 a 59	5	6
60 a 79	11	14
80 a 99	19	24
100 a 119	10	12
120 a 139	4	5
140 a 159	3	4
Total	80	100

Nota:

Por lo general no incluimos los datos perdidos en el cálculo de porcentajes.

La frecuencia de las respuestas en cada grupo se calcula como el porcentaje de elementos del estudio para los que se obtuvo datos (o, si se hizo una pregunta a los entrevistados, el porcentaje de aquellos que la contestaron).

Sin embargo, el número de datos perdidos (por ejemplo, la gente que no contestó una pregunta) es un indicio útil de qué tan adecuada fue la recolección de datos. Por lo tanto, este número debe mencionarse, por ejemplo, como una nota de la tabla. (Véase la tabla 22.2).

Recuerde que “no sabe” es una categoría especial que NO debe contarse como datos perdidos. Si corresponde, “no sabe” debe aparecer en la tabla como una categoría.

Se debe ser cauteloso al calcular e interpretar los porcentajes cuando el total es una cifra menor, porque una unidad de más o de menos va a establecer una gran diferencia en términos de porcentajes. Como regla general, los porcentajes no deben emplearse cuando el número total es inferior a 30.

Por consiguiente, se recomienda que el número de observaciones o el total de casos estudiados se dé siempre junto con el porcentaje.

2. Proporciones

Algunas veces las frecuencias relativas se expresan en proporciones en lugar de en porcentajes.

Una *proporción* es una expresión numérica que compara una parte de las unidades del estudio con el todo; una proporción puede expresarse como una *fracción* o en *decimales*.

Ejemplo 5:

De un total de 55 pacientes que asisten a una clínica un día en particular, 22 son hombres y 33 son mujeres. Podemos decir que la proporción de hombres es $22/55$ o $2/5$, lo que equivale a 0.40.

Tome en cuenta que cuando una proporción expresada en decimales se multiplica por 100, el valor que se obtiene es un porcentaje. En el ejemplo, 0.40 es equivalente a 40%.

3. Razones

Una *razón* es una expresión numérica que indica la relación en cantidad, dimensión o tamaño entre dos o más partes.

En el *ejemplo 5* que vimos un poco antes, la razón de hombres a mujeres es de 22:33, o 2:3.

4. Tasas

Una *tasa* es la cantidad, suma o grado de un evento o una enfermedad medido durante un determinado periodo.

Las tasas comúnmente usadas en el sector salud son:

- Tasa de natalidad = el número de nacidos vivos por cada 1,000 habitantes en un año
- Tasa de mortalidad = el número de defunciones por cada 1,000 habitantes en un año
- Tasa de mortalidad infantil (TMI) = el número de muertes de niños menores de un año por cada 1,000 nacimientos logrados en un año
- Tasa de mortalidad materna (TMM) = El número de defunciones maternas relacionadas con el embarazo en un año por cada 100,000 nacimientos logrados ese mismo año

- Tasa de incidencia = El número de personas que contraen una enfermedad por cada 1,000 o 100,000 habitantes expuestos en un periodo dado (por lo general un año)
- Tasa de prevalencia = El número de personas afectadas en una población en un periodo dado (por lo general un año)

IV. GRÁFICOS

Si su informe contiene muchas tablas descriptivas, puede ser más legible si presenta lo más importante en ilustraciones.

Las ilustraciones usadas con mayor frecuencia en la presentación de datos incluyen:

- Gráficos de barras
 - Gráficos de pastel
- } para datos categóricos
- Histogramas
 - Gráficos de línea
 - Diagramas de distribución
 - Mapas
- } para datos numéricos

A continuación se muestran ejemplos de estos gráficos.

1. Gráfico de barras

Los datos del *ejemplo 2* pueden presentarse en un gráfico de barras empleando ya sea frecuencias absolutas o frecuencias relativas/porcentajes (véase el gráfico 22.1).

Tome en cuenta que cuando presente los datos en porcentajes, debe indicar el tamaño de la muestra.

2. Gráfico de pastel

Un gráfico de pastel puede usarse para la misma serie de datos, para darle al lector una rápida perspectiva general de una manera diferente. Un gráfico de pastel ilustra la frecuencia relativa de varios temas. Todos los segmentos del gráfico de pastel deben sumar 100%.

Ilustración 22.1 Frecuencia relativa de la escasez de medicamentos contra la malaria en centros de salud rurales (n = 148)

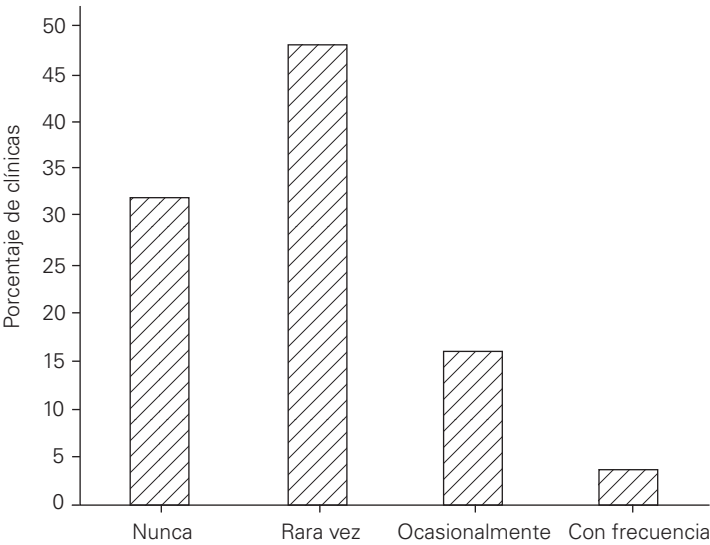
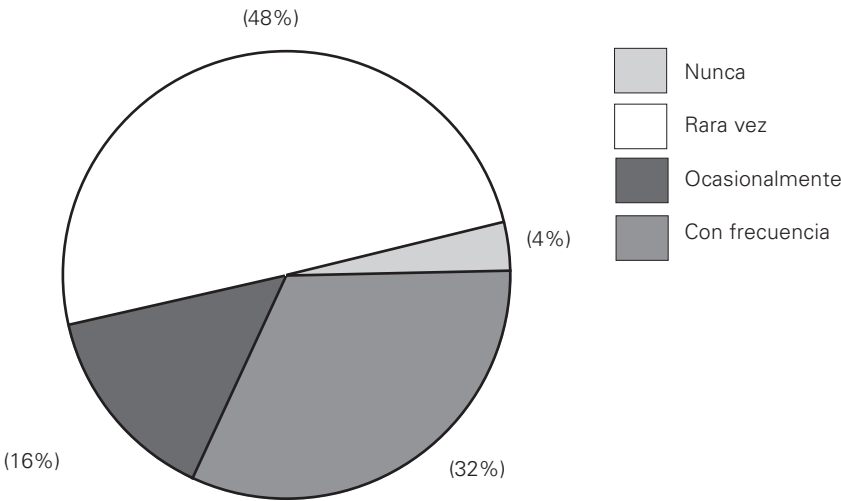


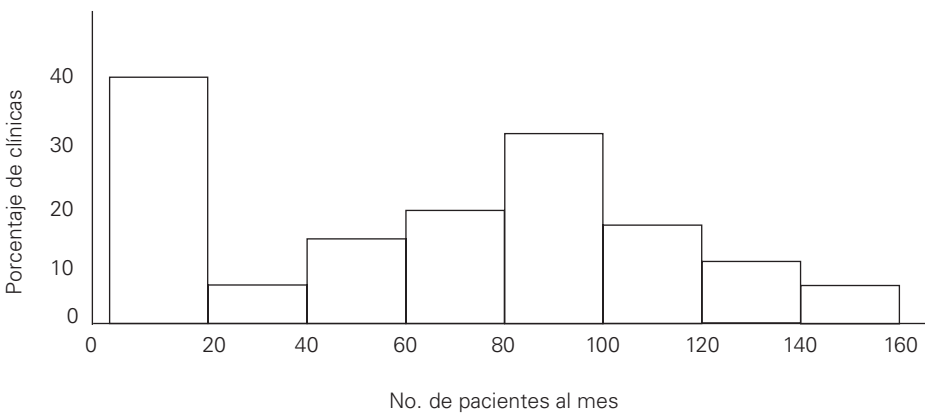
Ilustración 22.2 Frecuencia relativa de la escasez de medicamentos contra la malaria en centros de salud rurales (n=148)



3. Histogramas

Los datos numéricos se presentan muchas veces en histogramas, que son muy similares a los gráficos de barras utilizados para presentar datos categorizados. Una diferencia importante, sin embargo, es que las “barras” de un histograma están conectadas (a menos que haya una brecha entre los datos), mientras que en un gráfico de barras, éstas no lo están, puesto que cada categoría tiene un nombre distinto. En la *ilustración 22.3* se presentan en un histograma los datos del *ejemplo 4*.

Ilustración 22.3 Porcentaje de clínicas que tratan una cantidad diferente de pacientes de malaria al mes (n = 80)



4. Gráficos de línea

Un gráfico de línea es particularmente útil para presentar datos numéricos si se desea mostrar una tendencia en un lapso dado. Los datos del *ejemplo 3* pueden presentarse en un gráfico de línea tal como se hace en la *ilustración 22.4*.

Es fácil mostrar dos o más distribuciones en un gráfico, siempre y cuando la diferencia entre las líneas sea fácil de distinguir. Así, pueden compararse distribuciones de frecuencias de diferentes grupos como, por ejemplo, la distribución de la edad entre hombres y mujeres, o entre casos y controles.

5. Diagramas de distribución

Los diagramas de distribución son útiles para mostrar información de dos variables que puedan estar relacionadas. El ejemplo de un diagrama de distribución dado anteriormente se utiliza en el *módulo 31*, en donde tratamos los conceptos de asociación y correlación. (Ilustración 22.5)

Ilustración 22.4 Número diario de pacientes de malaria en los centros de salud del distrito X

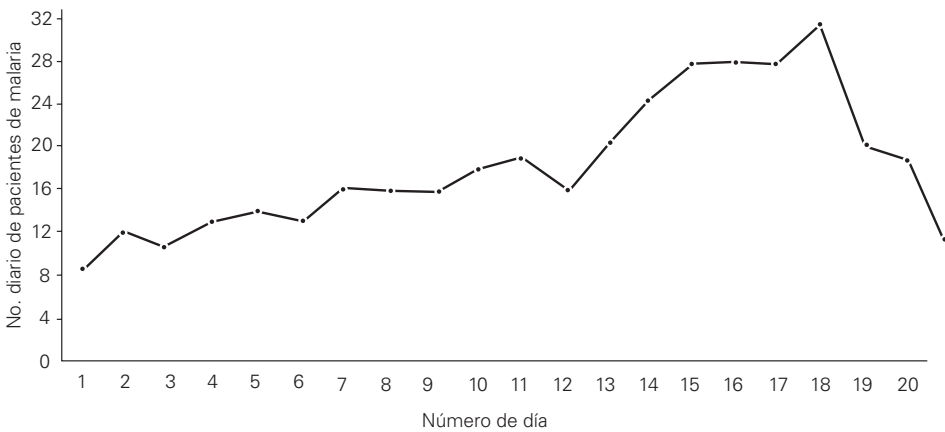
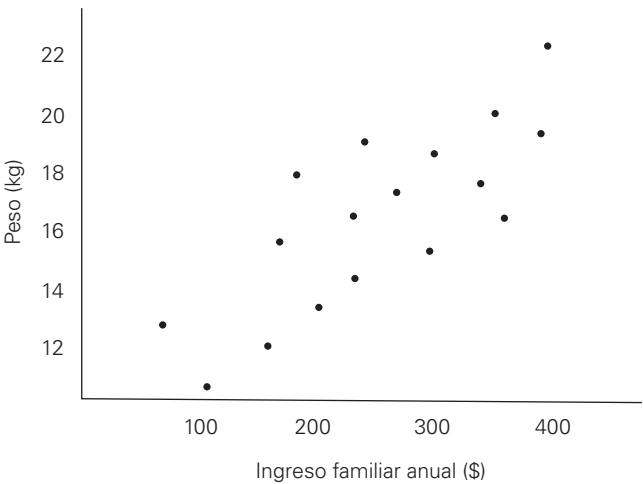


Ilustración 22.5 Peso de niños de 5 años de acuerdo con el ingreso familiar anual



Nota:

Todas las ilustraciones presentadas en su informe de investigación deben tener números y etiquetas y títulos claros.

Además de los gráficos comentados para presentar información, pueden usarse *mapas*. Puede mostrarse en un mapa, por ejemplo, el área en donde se llevó a cabo un estudio. Si éste exploró la epidemiología del cólera, podría hacerse un mapa que muestre la distribución geográfica de los brotes, o de las fuentes de agua protegidas, para mostrar que existe una asociación. Si el estudio es de la cobertura de vacunas, se podría hacer un mapa de los sitios en donde hay clínicas en cada población y la cobertura de vacunación entre los menores de 5 años, que muestre que la distancia del hogar a la clínica es un factor importante del estatus de la vacunación.

V. MEDIDAS DE TENDENCIA CENTRAL

La distribución de frecuencias y el histograma son medios útiles para ver series de observaciones de una variable. En muchos casos, son esenciales para entender los patrones de los datos. Sin embargo, si más adelante se quiere sintetizar una serie de observaciones, muchas veces conviene utilizar una medida que pueda expresarse con un solo número.

Antes que nada, queremos tener una medida para el centro de la distribución. Las tres medidas empleadas con este propósito son la media, la MEDIANA y la moda.

1. Media

La media (o media aritmética) es también conocida como el *promedio*. Se calcula sumando el total de los resultados de todas las observaciones y dividiéndolos entre el total de observaciones. Tome en cuenta que la media puede calcularse sólo para datos numéricos.

Ejemplo 6:

La medición de la estatura de siete niñas dio los siguientes resultados:

141, 141, 143, 144, 145, 146, 155 cm (un total de 1015 cm entre las 7 medidas)

La media es, entonces, $1015/7$, lo que da 145 cm.

2. Mediana

La *mediana* es un valor que divide una distribución en dos mitades iguales.

La mediana es útil cuando algunas de las medidas son mucho mayores o mucho menores que las demás. La media de semejantes datos se sesga hacia estos valores de los extremos; no nos da, por consiguiente, una buena medida del centro de la distribución. La mediana no se ve afectada por los valores de los extremos, su valor –llamado también valor central o intermedio–, se obtiene de la siguiente manera:

- Liste las observaciones en orden de magnitud (del valor más bajo al más alto, o viceversa)
- Cuente el número de observaciones (n)
- El valor de la mediana es el valor correspondiente al número de observaciones $(n + 1) / 2$, si n es impar, o el promedio de los dos números centrales

Ejemplo 8:

El peso de 7 mujeres embarazadas es: 40, 41, 42, 43, 44, 47, 72 kg.

El valor de la mediana es el valor correspondiente a la observación número: $(7 + 1) / 2$, que es el cuarto: 43 kg.

La media del peso de esta serie de observaciones es, en cambio, 47 kg. Este ejemplo ilustra cómo la media se ve afectada por los valores de los extremos (en este caso 72 kg), mientras que esto no ocurre con la mediana. Si el peso mayor de esta serie de observaciones hubiera sido 51 kg en vez de 72 kg, la mediana habría seguido siendo 43 kg, pero la media habría sido 44 kg.

Si hubiera 8 observaciones: 40, 41, 42, 43, 44, 47, 49 y 72, la mediana sería 43.5 kg (el promedio de 43 y 44); y la media sería 47.25 kg.

3. Moda

La *moda* es el valor que aparece con mayor frecuencia en una serie de observaciones.

La moda no es muy útil para datos numéricos continuos, pero sí lo es para datos numéricos agrupados.

En el *ejemplo 4* (número de pacientes tratados de malaria en una clínica) la moda es “0 a 19”, puesto que este dato se registra más frecuentemente (25 veces de 80).

La moda puede utilizarse también con datos categóricos, ya sean nominales u ordinales.

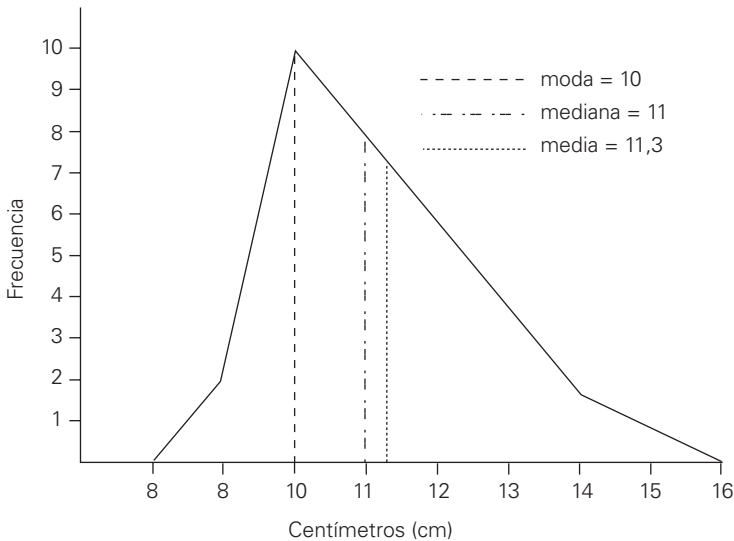
En el *ejemplo 1* (método de planificación familiar), la moda es “ninguno”. En el *ejemplo 2* (número de clínicas que se quedan cortas de medicamentos) la moda es “rara vez”.

En *síntesis*, la media, la mediana y la moda son todas medidas de tendencia central, de las que la media es la más empleada; contiene más información porque su cálculo toma en cuenta el valor de cada observación.

Sin embargo, la media se ve muy afectada por valores distantes del centro de la distribución, mientras que la mediana y la moda no. El cálculo de la media está al principio de procedimientos estadísticos más complejos para describir y analizar los datos.

La *ilustración 22.6* muestra una distribución curva en que la media, la mediana y la moda tienen diferentes valores.

Ilustración 22.6 Media, mediana y moda en una distribución curva



TRABAJO EN EQUIPO

1. Describa su(s) muestra(s) en términos de variables contextuales (sexo, edad, etc.) y variables dependientes (por ejemplo, desertor/ cumplidor, usuario/ no-usuario).
2. Cerciórese de haber contado las frecuencias de todas las variables de su estudio (de los datos de sus hojas maestras). Calcule los porcentajes en relación con el

total de unidades de estudio (o calcule proporciones/ razones/ tasas en donde corresponda).

3. Revise sus objetivos para determinar de qué variables necesita tablas de frecuencia para incluirlas en su informe. Por lo general, las tablas de frecuencia se presentan para algunas de las variables contextuales, la(s) variable(s) dependiente(s) y las variables independientes más importantes. Prepare las tablas de frecuencia.
4. Haga histogramas, gráficos de barras, gráficos de pastel y gráficos de línea, si le parecen de utilidad. Prepare descripciones breves para interpretar lo que significa cada una de las ilustraciones empleadas.
5. Calcule medias, medianas y modas, si aplica, e interprete los resultados.
6. Familiarícese con los resultados y trate de entender lo que significan.

Referencias

Todos los libros de texto de epidemiología y estadística mencionados en los módulos 9 y 28.

Notas del instructor

Módulo 22 DESCRIPCIÓN DE LAS VARIABLES

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
3 horas+	Trabajo en equipo

Introducción y discusión

- Es probable que los participantes estén familiarizados con algunos de los conceptos presentados en este módulo, tales como los porcentajes y las proporciones. Puesto que en esta etapa los grupos ya deben haber preparado distribuciones de frecuencia (incluyendo el cálculo de porcentajes), en la presentación debe usted mencionar estos conceptos sólo brevemente—para no perder su interés—, sobre todo si el nivel de conocimiento de los participantes es alto. Sin embargo, debe dar una atención especial a lo que habrá que hacer con los datos perdidos al calcular los porcentajes.

- Aunque en el módulo se definen los conceptos de porcentaje, proporción, razón y tasa, es más importante dar ejemplos o pedirles a los participantes que los den ellos mismos.
- Al presentar el *ejemplo 3*, pídales a los participantes que describan cómo agruparon sus datos numéricos, y revise con ellos si establecieron demasiadas o muy pocas categorías.
- Los ejemplos no sólo deben revisarse, sino también usarse en ejercicios informales. Por ejemplo, pregúnteles a los participantes cuáles son la media, la mediana y la moda de una serie dada de medidas, en vez de darles las respuestas.

Trabajo en equipo

- Antes de que el grupo calcule la frecuencia de las variables de las hojas maestras, debe revisar si categorizó correctamente los datos. Cerciórese también de que hayan definido el total de informantes (unidades de estudio) de cada grupo estudiado.
- Recuérdele al grupo que sólo debe hacer las tablas de frecuencias de las variables que va a describir en el informe final. Por lo general, son necesarias las tablas de algunas de las variables contextuales del (los) grupo(s), y a veces también de las variables independientes más importantes. Varias de las demás variables contextuales y las variables independientes deben presentarse en tabulaciones cruzadas (*módulo 24*).

Módulo 23

ANÁLISIS DE LOS DATOS CUALITATIVOS

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 23 DANÁLISIS DE DATOS CUALITATIVOS

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Describir* maneras eficientes de ordenar y sintetizar datos cualitativos.
2. *Indicar* por qué es esencial comenzar la síntesis y el análisis durante el trabajo de campo.
3. *Listar* los pasos más importantes del análisis de los datos cualitativos y formular conclusiones.
4. *Hacer* un esquema de cómo va a ordenar y a sintetizar los datos cualitativos, y a hacer el análisis subsecuente.
5. *Planificar* cómo exponer sus datos cualitativos para que se integren de manera efectiva con el resto de los datos.
6. *Indicar*, enseguida o al terminar el análisis de los datos, qué otras actividades va a realizar para comprobar o confirmar sus hallazgos y probar, así, *su validez*.

I. Introducción.

II. Procedimientos para procesar y presentar datos cualitativos.

III. Formulación y verificación de las conclusiones utilizando distintas series de datos.

IV. Informe de los datos.

V. Estrategias posteriores para comprobar o confirmar los hallazgos, y probar su validez (opcional).

I. INTRODUCCIÓN

En *módulos* anteriores (9, 10, 13), se señaló que las técnicas de investigación cualitativa se usan para comprender profundamente determinadas situaciones o problemas de los que sabemos poco. Las técnicas cualitativas –tales como el uso de *entrevistas de estructura flexible con preguntas abiertas, grupos de discusión, observaciones o abordajes proyectivos y participativos*– van a ser, por tanto, adecuadas en muchos estudios, especialmente en el arranque. Pueden, además, ser las únicas técnicas confiables para temas delicados.

Independientemente de cómo y para qué se recogieron los datos, el investigador termina, por lo general, con un buen número de páginas de *texto escrito* que debe analizar.

Aunque los procedimientos y los resultados del análisis de los datos cualitativos difieren de los del análisis de los datos cuantitativos, los principios no son tan diferentes. En ambos casos el investigador debe:

- describir a la población de la muestra
- ordenar y reducir o codificar los datos (procesamiento de datos)
- presentar una síntesis de los datos de manera que sean fáciles de interpretar, por ejemplo, preparando hojas de compilación, rota-folios, diagramas o matrices
- formular conclusiones, relacionarlas con otras series de datos del estudio y decidir cómo integrar los datos en el informe
- idear, de ser necesario, estrategias para comprobar o confirmar los datos (cualitativos) a fin de probar su validez

Examinemos ahora más detalladamente cada uno de estos puntos

II. PROCEDIMIENTOS PARA PROCESAR Y PRESENTAR DATOS CUALITATIVOS

1. Descripción de la población de la muestra en relación con los procedimientos de muestreo

Un primer paso útil para procesar los datos (así como para informar de los hallazgos) es describir a los informantes. Si los números lo permiten, pueden tabularse datos contextuales relevantes como, por ejemplo, la edad, el sexo, la ocupación, la escolaridad o el estado civil, tal como se hace en los estudios cuantitativos.

Sin embargo, como los datos cualitativos se obtienen de muestras reducidas (a veces unos cuantos informantes clave, algunos grupos de discusión o varias

observaciones) es necesaria una mayor información para ubicar los datos en su contexto.

Por ejemplo, ¿quiénes fueron los informantes clave?, ¿qué lo llevó a optar por ellos? ¿Quiénes tomaron parte en los grupos de discusión? ¿Cómo son los participantes de los grupos seleccionados y qué tan representativos son de la población de estudio? En cuanto a las observaciones, ¿en qué circunstancias se llevaron a cabo? ¿Quiénes fueron observados y por quién?

Si no se proporciona esta información, la interpretación de los datos podría dar la impresión de ser aleatoria.

2. Clasificación y codificación de los datos

Vamos a comentar dos tipos de datos cualitativos:

- las respuestas a las preguntas abiertas
- el relato más elaborado de las entrevistas semi-estructuradas o de los GD.

1) Respuestas a las preguntas abiertas

Los datos cualitativos más comúnmente recogidos son las respuestas a preguntas abiertas; éstas forman parte de todo estudio de ISS. Cuando elaboró usted su protocolo en el *módulo 13*, hizo un ejercicio de clasificación sistemática de dichos datos. Para analizar los diferentes pasos,¹ vamos a profundizar en las respuestas a la pregunta “¿Por qué fuma usted?”

- (1) El primer paso en el análisis es *listar* las respuestas de una *muestra de 20-25 informantes* tal y como se obtuvieron (añadiendo el número del cuestionario para no perder la conexión con los demás datos del mismo informante).
- (2) Enseguida *lea* cuidadosamente las respuestas, recordando el propósito de la pregunta. Se espera que la pregunta “¿por qué está usted fumando?” ayude a los estudiantes de enfermería a idear una intervención contra el cigarro.
- (3) *Establezca categorías gruesas* de respuestas que puedan ir juntas y *codifíquelas* con una palabra clave. Por ejemplo, las respuestas 3 (Me brinda placer) y 14 (Me gusta hacer aros de humo) podrían etiquetarse con el término “placer”, que podría abreviarse con el código *plc*.
- (4) *Enliste* nuevamente todas las respuestas, pero ahora *por código*, de manera que obtenga unas 5-7 listas breves, por ejemplo:

¹ Estos pasos se adaptaron de Willms y Johnson (1996).

<i>Placer (plc)</i>	<i>Ser sociable (soc)</i>	<i>Dar confianza (cfz)</i>
2. Me gusta la sensación del cigarro en mi mano	10. Todos mis amigos son fumadores	6. Me da confianza y siento que domino la situación cuando fumo
3. Porque me da placer	11. Me ayuda a hacer sentir a la gente más en confianza y más cómoda cuando les ofrezco un cigarro	7. Me ayuda a pensar mejor
5. Me gusta hacer aros de humo con la nariz y la boca		18. Me ayuda a reducir la presión y la tensión en el trabajo
14. Me gusta soplar el humo		17. Me ayuda a relajarme
15. Me gusta el sabor		

(5) *Interprete* cada una de las listas y termine con unas 5-7 categorías significativas, encabezadas con una característica como palabra clave. Por ejemplo: *Placer, ser sociable, estatus, confianza, adicción, rebeldía*. Puede comentar la necesidad de dividir algunas de las categorías o de combinar las que tengan pocas respuestas. Las respuestas 17 y 18, por ejemplo, podrían separarse en la categoría *reducir la tensión*. En ese caso habría siete categorías. La categoría *rebeldía* podría tener dos respuestas: 4 ¡No veo por qué tendría que dejar de fumar!, y 12 ¿Y por qué no! Los puntos de exclamación indican que lo que subyace tras la motivación de la respuesta es rebeldía, más que desconocimiento. Si el entrevistador no los hubiera agregado, habría sido difícil codificar estas respuestas.

Haga ahora una interpretación tentativa de acuerdo con la disposición que hayan mostrado sus informantes para cambiar su comportamiento. Para quienes fuman por *placer* o para *socializar* podría ser más fácil dejar de fumar. Los que son *adictos* pero trataron de dejarlo y quienes sienten que fumar les da un *estatus* podrían formar una categoría intermedia. A quienes fuman para afirmar su *confianza en sí mismos* o para *bajar la tensión*, así como a los que se muestran *rebeldes* ante la pregunta acerca de por qué fuman, podría serles más difícil dejar el cigarro.

- (6) Pruebe enseguida un nuevo lote de 20-25 respuestas y *revise si las etiquetas funcionan*. Es muy posible que en esta etapa todavía cambie algunas etiquetas o decida agregar nuevas categorías o hacer nuevas combinaciones.
- (7) *Haga una lista final* de categorías etiquetadas y codifique todos los datos, incluyendo los que acaba de procesar, con códigos abreviados.

Revise ahora si debe apegarse a su interpretación tentativa de los datos, y lo que eso significa respecto del contenido de los mensajes para tratar las diferentes razones de fumar. Este *análisis del contenido* es el principal propósito del análisis. Sin embargo, al *contar* las respuestas de cada etiqueta, también va a saber qué tan comunes son estas distintas razones.

2) Relato elaborado

Los datos de las entrevistas con informantes clave o de los grupos de discusión (GD) son, por regla general, más voluminosos que las respuestas a las preguntas abiertas. Las notas de campo y las cintas grabadas y transcritas al pie de la letra (véanse el *módulo 10C* sobre los GD y el *módulo 13*), pueden dar lugar a páginas y páginas de narración. Al analizar los textos, descubrimos, por lo general, que no importa que tan buenas hayan sido nuestras instrucciones para la discusión, los datos contienen información valiosa pero también una serie de minucias que no son esenciales. Además, los datos no suelen estar presentados en el orden en que los necesitamos para nuestro análisis, pues los informantes pueden saltar de un tema a otro.

Para facilitar el análisis, debemos *ordenar y reducir* los datos. Es mejor ordenarlos en relación con los objetivos y los temas de discusión. Aquí también conviene seguir sistemáticamente varios pasos.

(1) *Relea* sus objetivos y los temas de discusión.

(2) *Lea* cuidadosamente un cierto número de las entrevistas, los registros de los GD y las observaciones que quiera procesar. Numere el material de acuerdo con el tema general de discusión al que correspondan. Utilice un marcador amarillo para destacar los comentarios particularmente ilustrativos. Utilice los márgenes para definir sub-temas.

Por ejemplo, en un estudio de género y lepra llevado a cabo en diferentes países (usado como ejemplo en los *módulos 4, 8 y 11*) parecía que el estigma del tema de discusión debía diferenciarse de acuerdo con el diferente escenario social en que ocurría: entre los parientes cercanos (padres-hijos), el cónyuge, los suegros y los miembros de la comunidad. Más aún, se tuvo que distinguir entre la auto-estigmatización (por ejemplo, una mujer a quien se le diagnosticó lepra animaba a su esposo a casarse por segunda vez para evitar el divorcio, o un paciente que no asistía a las reuniones de la comunidad por temor a que lo evadieran) y la estigmatización de los demás. También pudieron distinguirse diferentes grados de estigmatización, que iban del rechazo leve a la total expulsión. Si el estigma fuera el tema (11) en su lista de discusión, tendría que marcar todo lo relacionado con el estigma con un (11) en el margen, y agregar, también en el margen,

palabras clave tales como *auto-estigma*, *cónyuge*, *suegros*, *comunidad*, así como palabras clave tales como *duerme(n) sep(arados)* o *divorciados* para indicar la gravedad del estigma. (Vea un ejemplo en el *anexo 10C.2* del módulo 10C).

- (3) *Enliste todas las palabras clave que correspondan a un tema determinado en las sub-categorías* que se hayan creado en (2). Por ejemplo, todo lo que corresponda al estigma podría subdividirse y enlistarse en los cuatro escenarios sociales principales en que se vio que el estigma se manifiesta.
- (4) *Interprete* los datos; por ejemplo, distinga las principales formas en que el estigma se manifiesta en estos distintos escenarios sociales, procure establecer un orden clasificatorio de la gravedad y vincúlelo con otras variables (tales como el grado de deformidad, el estatus socio-económico) a fin de entender las diferencias en el estigma.
- (5) *Codifique entonces todos sus datos cualitativos* de esta manera. De ser necesario, adapte su esquema de codificación conforme ordena, codifica e interpreta más datos. En ese caso, debe usted leer de nuevo y posiblemente re-codificar el material que ya había procesado.

Nota:

Es posible que usted haya analizado y codificado sus datos cualitativos en el campo (como se recomendó en el *módulo 13*), para ajustar sus guías de las entrevistas o su lista de temas. En dado caso, puede elaborar su lista final de códigos en un solo ciclo y no en dos.

No obstante, en vez de elaborar un sistema de codificación muy pormenorizado para sus datos crudos, puede usted refinar su interpretación conforme registre sus datos codificados en grueso y sintetizados en *hojas de compilación*.

3. Síntesis de los datos en hojas de compilación

Tras ordenar los datos, debe sintetizarlos. Un primer paso útil es sintetizar en hojas de compilación separadas todos los datos de cada una de las unidades de estudio por población de estudio.

Al igual que las hojas maestras para los datos cuantitativos, las hojas de compilación para los datos cualitativos consisten en un determinado número de columnas encabezadas con los temas cubiertos por el estudio. Éstas todavía pueden subdividirse en los temas, más reducidos, que haya usted identificado y codificado cuando ordenó los datos (véase el anexo 23.1). Se asigna un número a cada entrevista, GD u observación, que se captura consecutivamente en esa

secuencia en la respectiva hoja de compilación. Si hubiera diferentes categorías de informantes en una misma población de estudio como, por ejemplo, madres jóvenes y madres mayores, o pacientes masculinos y pacientes femeninas, los datos de estos grupos deben ingresarse en hojas por separado. Si los temas cubiertos en esos sub-grupos no son idénticos, es importante ser sistemáticos y seguir a grandes rasgos la misma secuencia temática para cada una de las categorías de informantes. La información capturada se sintetiza en palabras y oraciones clave lo bastante claras como para recordar los planteamientos de los informantes. (En la medida en que el número de cada unidad de estudio se captura en la hoja de compilación, siempre se puede regresar a los datos originales y presentar el planteamiento completo, por ejemplo en una presentación o en el informe de investigación).

Ahora tiene usted un panorama general de todos los datos por población de estudio en una o varias hojas grandes. Si lee usted las columnas, tiene una lista de las respuestas de todos los miembros del grupo sobre un determinado (sub-)tema. Si lee las hojas horizontalmente, puede relacionar –por informante– diferentes temas entre sí o a las características personales del informante. También es fácil comparar las respuestas de los diferentes grupos sobre temas específicos, contrastando simplemente las hojas de compilación.

Por ejemplo, en el *anexo 23.1* se presentan los datos personales de los pacientes con lepra (recién dados de alta) y un cierto número de temas y sub-temas comentados con ellos. El estigma vivido realmente, que era al principio un tema, se subdividió en la hoja de compilación en los cuatro principales escenarios sociales en que éste puede darse: parientes sanguíneos cercanos, matrimonio, el círculo más amplio de la familia política, y la comunidad. En cada uno de estos grupos todavía pueden hacerse distinciones más finas (por ejemplo, la comunidad puede consistir en vecinos, amigos, colegas, condiscípulos o miembros de la comunidad más distantes). Si las muestras son reducidas, éstos pueden ingresarse todos bajo el encabezado “comunidad”. Pueden agregarse códigos (en *italicas*) a los planteamientos presentados en palabras clave, por ejemplo *gran temor y preocupación* bajo el encabezado “primera reacción”. De los tres ejemplos presentados, enseguida se hace evidente (y lo confirma el análisis del total de los datos en los cuatro países) que en general el temor al estigma que surge cuando los pacientes oyen el diagnóstico de la lepra es mayor que el estigma que sufren realmente. El paciente (12) es, al respecto, una excepción. Irónicamente, cuando a ella la dieron de alta, el esposo que se divorció de ella había muerto de otra enfermedad. La comparación horizontal de los datos del paciente (1) nos muestra que es muy poco probable que los amigos de este hombre no sepan de la enfermedad, puesto que incluso después de haber sido dado de alta tenía signos visibles. Aquí los investigadores tuvieron que entrevistar a los amigos para enterarse de si efectivamente el hombre había sido (o no había sido) estigmatizado por la comunidad.

Puede notar que la interpretación y el etiquetado de los datos se facilitan con el uso de las hojas de compilación, puesto que el investigador puede visualizar todos los aspectos de sus informantes incluso si ve un solo aspecto a la vez de toda la población de estudio.

Un siguiente paso en la síntesis puede ser la combinación, el contraste o el análisis posterior de temas importantes en presentaciones gráficas tales como las matrices, los diagramas, los diagramas de flujo y las tablas.

4. Síntesis adicional de los datos en matrices, gráficos y tablas

Matrices

Las matrices pueden emplearse para comparar tanto los datos cuantitativos como los datos cualitativos. Respecto de los datos cualitativos, podríamos comparar diferentes grupos o series de datos de variables importantes, presentados en palabras clave.

Una *matriz* es un gráfico semejante a una tabulación cruzada, pero que contiene palabras (así como, a veces, números).

En un GD sobre el cambio de prácticas de destete, los investigadores listaron las respuestas de las madres jóvenes y las de las madres mayores de 45 años sobre la introducción de comida blanda para lactantes. Luego sintetizaron estas respuestas en un matriz:

Ilustración 23.1 Matriz sobre la introducción de comida blanda para lactantes entre madres de diferentes grupos de edad

Grupos de edad	Inicio de comida blanda	Tipo de comida	Frecuencia de cb/al día
Madres jóvenes (20 a 30 años)	Rango: 4 a 7 meses Promedio: 6 meses	<ul style="list-style-type: none">• Avena blanda• Avena blanda con cacahuete machacado• Puré de papas, fruta machacada, galletas remojadas	2 a 4 veces al día <ul style="list-style-type: none">• Dependiendo de la disponibilidad de la madre y la nana• Dependiendo del apetito del niño
Madres mayores de 45 años	Rango: 5 a 11 meses Promedio: 8.5 meses	<ul style="list-style-type: none">• Avena blanda• Fruta blanda	1 a 2 veces al día <ul style="list-style-type: none">• Dependiendo de la disponibilidad de la madre y la nana• Dependiendo del apetito del niño

Este tipo de presentación facilitó a los investigadores concluir que:

- las madres más jóvenes comienzan a dar comida blanda, en promedio, 2.5 meses antes que la generación de sus propias madres
- las madres más jóvenes usan una mayor variedad de comida blanda para el destete que las mujeres de generaciones precedentes
- las madres más jóvenes dan comida blanda a sus lactantes con mayor frecuencia, aunque por las mismas razones que lo hicieron sus madres

Las matrices facilitan considerablemente el análisis de los datos. Son la forma más común de presentación gráfica de los datos cualitativos. Pueden emplearse para ordenar y comparar información de diversas maneras, por ejemplo, de acuerdo con:

- la secuencia temporal (de procedimientos investigados en diferentes periodos, por ejemplo)
- el tipo de informantes (como en el ejemplo anterior)
- la ubicación geográfica de las fuentes de datos (para visualizar las diferencias entre poblaciones rurales y urbanas)

Diagramas

Un *diagrama* es una ilustración con cajas que contienen variables y flechas que indican las relaciones entre dichas variables.

Al analizar los problemas que querían investigar durante la elaboración de los protocolos, la mayoría de los grupos hicieron un diagrama. De manera similar, pueden hacerse diagramas para sintetizar los hallazgos de un estudio. (Véanse las *ilustraciones 23.2 y 23.3*).

Usted puede usar un diagrama para ilustrar un tema crucial en su estudio, combinando todos los datos cualitativos y cuantitativos disponibles.

Los diagramas, al igual que las matrices, pueden ser de gran ayuda para hacerse de una perspectiva general de los datos recogidos y orientar su análisis.

Diagramas de flujo

Los *diagramas de flujo* son un tipo especial de diagramas que expresan la secuencia lógica de una serie de acciones o decisiones.

Ilustración 23.2 Razones de la introducción prematura de comida blanda por las madres jóvenes

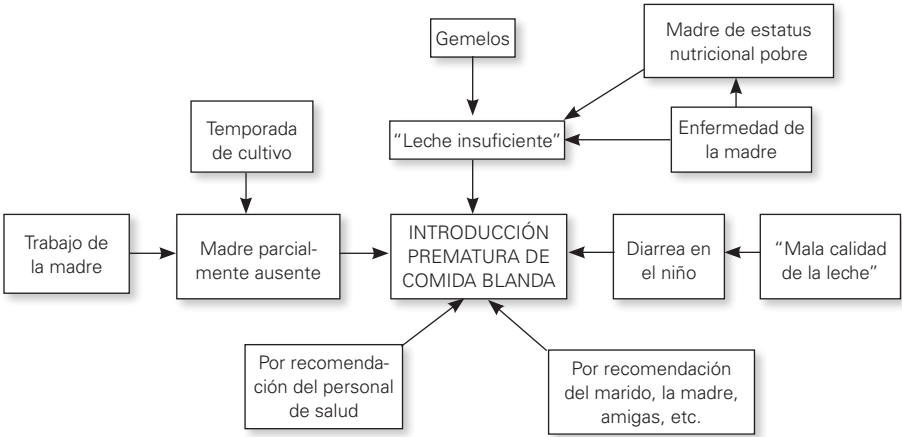
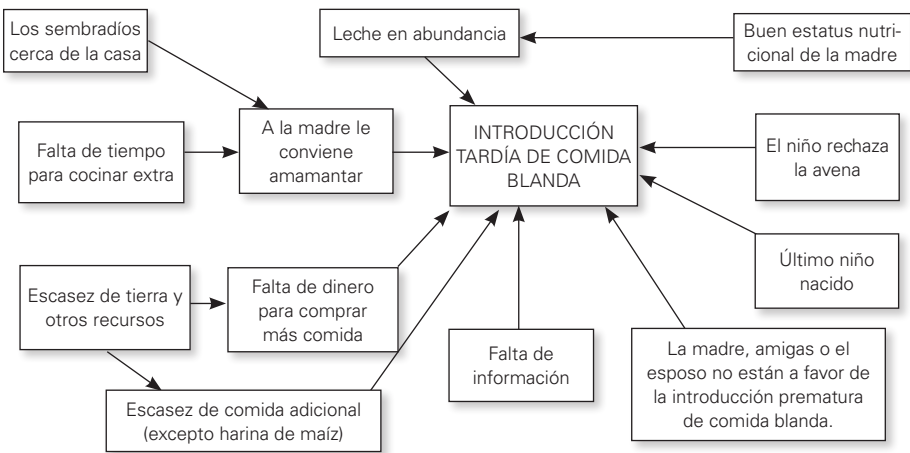


Ilustración 23.3 Razones de la introducción tardía de alimentos blandos por las madres jóvenes



La ilustración que precede a cada uno de los *módulos 1-18* e indica los pasos sucesivos de la elaboración del protocolo es un ejemplo de un diagrama de flujo.

Los diagramas de flujo son especialmente útiles para sintetizar diferentes flujos de acontecimientos interconectados. Por ejemplo, un equipo de asesoramiento en Bulawayo, Zimbabue, que entrevistó a profundidad a unas 95 personas con VIH positivo a lo largo de dos años, sintetizó alrededor de 100 páginas de material de entrevistas por informante, trazando cinco líneas (véase la ilustración 23.4). Una línea central presenta el desarrollo de la enfermedad a lo largo del tiempo, con crisis y periodos de relativo bienestar. Otra línea presenta diferentes formas de vigilancia médica; una

tercera línea, los flujos en el estatus económico relacionado con la enfermedad (por ejemplo, la pérdida del trabajo, la búsqueda de empleo en otra parte). Una cuarta línea presenta los posibles cambios en el estatus social, tales como el divorcio o el volver a casarse, mientras que una quinta línea presenta el estatus emocional del paciente, vinculado con los acontecimientos ocurridos en los cuatro campos mencionados (por ejemplo, el desenvolvimiento positivo, la depresión). Estos diagramas de flujo fueron sumamente útiles para comparar los datos por informante y entre diferentes grupos de informantes (por ejemplo, hombres/ mujeres, solteros/ casados). Permitieron destacar los efectos de la enfermedad en la vida de los diferentes grupos de pacientes y su manera de enfrentar la enfermedad.²

Tablas

Una *tabla* es un gráfico con filas y columnas, que tiene números en varias celdas o cajas.

Los datos cualitativos pueden también categorizarse, codificarse, insertarse en hojas maestras o en la computadora y *contarse*, junto con otros datos cuantitativos, y presentarse en tablas. Las respuestas a las preguntas abiertas de los cuestionarios se categorizan y sintetizan de esta manera. Aunque primero se *analiza el contenido* de las respuestas individuales en cada categoría. (Véanse la sección II-2 y la sección III de este módulo.)

III. FORMULACIÓN Y VERIFICACIÓN DE LAS CONCLUSIONES

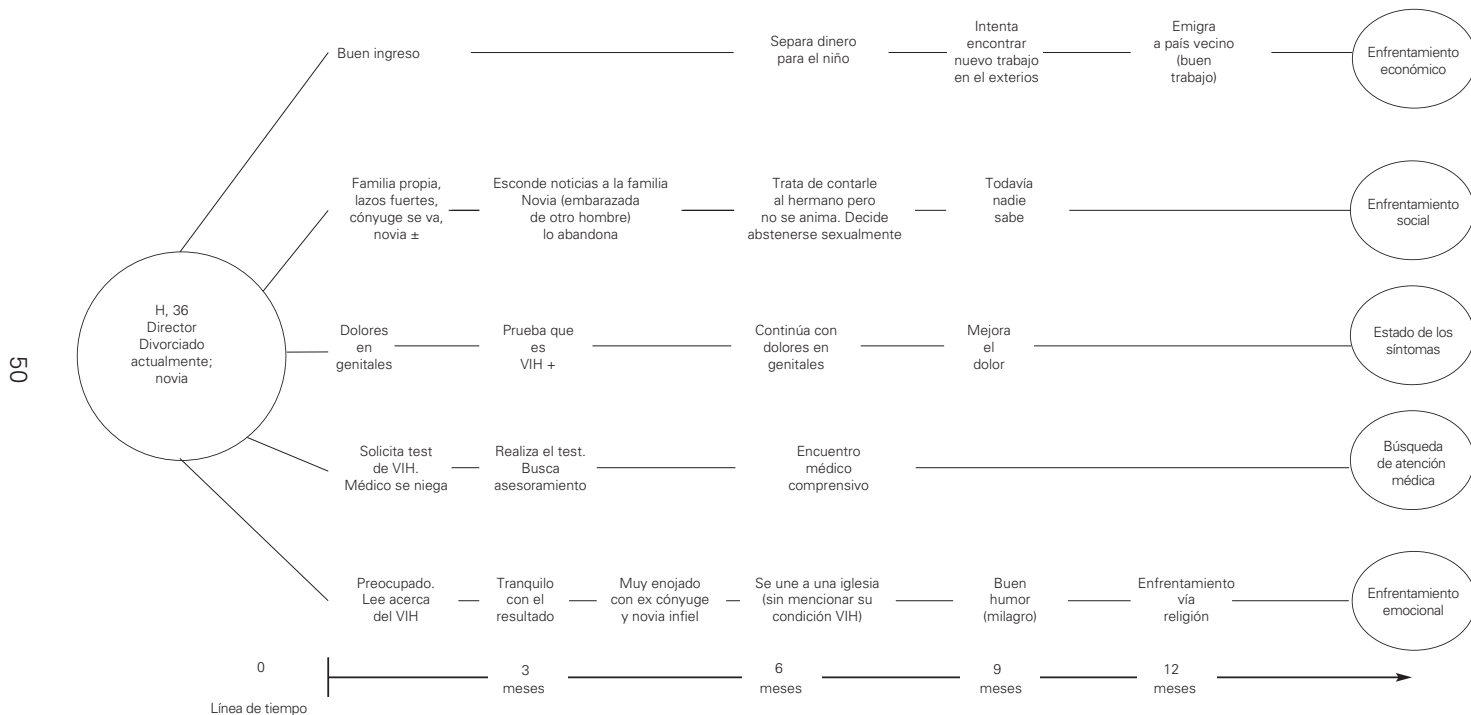
La formulación y la verificación de las conclusiones es la esencia del análisis de los datos. No obstante, no se trata de una actividad aislada. Cuando comenzamos a sintetizar nuestros datos en hojas de compilación, diagramas de flujo, matrices o diagramas, esbozamos continuamente conclusiones y, conforme avanzamos, modificamos o rechazamos algunas de ellas. Escribir ayuda también a generar nuevas ideas. Por consiguiente, *la redacción debe comenzar* desde el inicio del procesamiento y el análisis de los datos, aunque sea sólo para nosotros mismos. ¡No debe perderse ninguna reflexión creativa!

Nota:

La recolección, el procesamiento, el análisis y el informe de los datos cualitativos están estrechamente interrelacionados; no son (como en el caso de los datos cuantitativos) distintos pasos consecutivos. Con frecuencia resulta necesario volver a las notas de campo originales y verificar las conclusiones, recoger datos adicionales, si es que los datos disponibles resultan controvertidos, y obtener la opinión de todos los sectores interesados.

2 Meursing (1997) *A world of silence*.

Ilustración 23.4 Diagrama de flujo sobre la manera en que personas con VIH+ enfrentan su situación a lo largo del tiempo



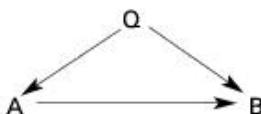
Identificación de las variables y sus asociaciones

En el *módulo 8* planteamos que a veces no sabemos lo suficiente de una situación como para definir anticipadamente las variables. Ciertas variables podrán definirse y asociarse a otras variables sólo durante el estudio, o al final de éste; no cabe al respecto el propósito *a priori* de *medirlas*. Para identificar estas variables, muchos estudios de ISS tienen partes cualitativas tales como preguntas abiertas, entrevistas con informantes clave, grupos de discusión u observaciones. El investigador que emplea semejantes aproximaciones cualitativas debe ser como un detective en busca de evidencia, que da cuenta de la evidencia contraria, y verifica los hallazgos buscando evidencia de apoyo independiente, hasta estar seguro de las posibles asociaciones entre determinadas variables, que arrojen luz sobre el problema que está siendo investigado.

Por ejemplo, si nos encontramos con que un cierto número de madres que destetan prematuramente a sus hijos trabajan, podemos pensar que tener un trabajo contribuye al destete precoz. Estudios similares hechos en otros lados con hallazgos similares apoyan esta conclusión (evidencia independiente). Sin embargo, podemos estar más seguros de que nuestra conclusión es verdadera sólo si entre las mujeres que destetan a sus hijos tardíamente hay muy pocas que trabajen, y para cada una de esas excepciones debemos buscar una explicación. ¿Las madres llevan consigo a los niños (al lugar de trabajo) o trabajan cerca de sus casas de modo que pueden alimentar a sus niños en los descansos? ¿O logran combinar exitosamente la lactancia con otras alternativas? De ser así, ¿por qué no más madres prueban esta combinación?, etc.

Búsqueda de variables de confusión o variables intermedias

A veces las variables parecen estar relacionadas, pero la asociación puede no explicarse fácilmente. Otras veces parece que, por lógica, las variables deberían ir juntas, pero no se puede encontrar una relación. En tales casos podría haber otra variable ("Q") influyendo en la asociación entre las dos variables implicadas, en cuyo caso ésta debe identificarse (véanse los módulos 8, 9 y 26).



Por ejemplo, uno esperaría que hubiera una relación entre la calidad del agua potable y la incidencia de diarrea. Se considera que la incidencia de diarrea debería decrecer conforme aumenta el número de grifos en el pueblo. Si no hay cambios en el transcurso del tiempo, podría haber una variable de confusión. A la gente, por ejemplo, podría disgustarle tanto el sabor del agua de grifo que podría usarla para todo excepto para beberla.

Nota:

Estas asociaciones inexplicables pueden surgir en cualquier estudio. La característica esencial de una aproximación cualitativa en la investigación es que busca deliberadamente asociaciones *durante* el trabajo de campo, de modo que puedan crearse las preguntas y herramientas adicionales que alumbren dichas relaciones. En la investigación cuantitativa, en cambio, cuando se busca medir objetivamente la fuerza de una asociación *prevista* entre dos variables, no deben cambiarse las herramientas una vez que el trabajo de campo se ha puesto en marcha.

Integración de los datos cualitativos y los datos cuantitativos

Hasta aquí hemos comentado el análisis de los datos cualitativos como una actividad por separado. Sin embargo, si un equipo de investigación ha recogido tanto datos cualitativos como cuantitativos, como sucede en la mayoría de los estudios de ISS, sería absurdo no verlos en combinación, puesto que ello puede inspirar un análisis más profundo y gratificante.

Por ejemplo, el equipo de la investigación indonesia “género y lepra” encontró, al analizar los datos de 4 500 nuevos pacientes con lepra registrados en los últimos cinco años, que la razón H/M era desfavorable en el grupo de edad de los 15 a los 44 años. Éste fue un descubrimiento desconcertante, puesto que las mujeres de este grupo de edad en Nepal estaban informando estar mucho mejor (aunque no tanto como los hombres). Entrevistas a profundidad al personal revelaron la sospecha de que las adolescentes y las jóvenes ocultaban sus parches de la piel por su asociación vergonzosa con lo sucio y lo feo. Esto brindó el incentivo para desglosar aún más los datos cuantitativos, lo que evidenció que la diferencia H/M al informar era todavía más pronunciada en el grupo de edad de 15 a 34 años y se nivelaba después de los 35. Las razones de esta diferencia de sexo relativamente amplia en los grupos más jóvenes se exploraron entonces todavía más.

Análisis del contenido de los datos cualitativos para propiciar la acción

Los datos cuantitativos sirven, en primer lugar, para convencer a las autoridades de la salud de *que* hay en efecto un grave problema mensurable; los datos cualitativos aportan ideas sobre *cómo* resolverlo. Los GD sobre la alimentación durante el destete entre madres jóvenes y madres que habían rebasado los 45 años, por ejemplo, dan lugar a muchas sugerencias sobre cómo hacer intervenciones con las mamás, que ellas consideren útiles y puedan realizar. De la misma manera, las entrevistas en profundidad con pacientes leprosos y ex-leprosos arrojan nuevas luces respecto de la mejor manera de aconsejar a los nuevos pacientes, a sus familiares cercanos y al/la cónyuge para reducir un temor innecesario.

Análisis de los datos cualitativos en la computadora

En los últimos tiempos se han refinado los recursos para el análisis de los datos cualitativos en la computadora. El *software* para el manejo de estos datos incluye la posibilidad de probar asociaciones estadísticas. Como en la ISS las cifras son por lo regular pequeñas y el análisis de contenido –que puede hacerse de manera manual–, es muy probablemente más importante que probar asociaciones, no vamos a desarrollar aquí estas técnicas. Más bien, vamos a remitir a los estudiantes interesados en el tema a los departamentos de Antropología y Psicología de las universidades que tengan experiencia en el uso de programas tales como *Qualitan* o *SPSS* para el procesamiento de datos cualitativos.

IV. INFORME DE LOS DATOS

Hay básicamente dos maneras de informar de los datos cualitativos; ambas forman parte de un estudio en que se emplearon diferentes técnicas de investigación. Una consiste en sintetizar los principales resultados cualitativos en una sección de hallazgos por separado, con ejemplos y citas acordes con los objetivos que dirigieron la recolección de estos datos en particular. Los resultados se comentan en el capítulo “Comentarios”; junto con los resultados de otras herramientas de recolección de datos cuantitativos, y se reflejan subsecuentemente, en el resumen de los hallazgos y las recomendaciones.

Otra posibilidad es integrar las diferentes series de datos en el capítulo de los hallazgos, ordenándolas de acuerdo con los objetivos del estudio completo. Si los datos cuantitativos y los cualitativos se analizaron –e incluso se recogieron– de manera integral, sería lógico presentarlos en un esquema también integral. Debe ponerse especial cuidado en no perder datos valiosos. Por consiguiente, se debe hacer –en cualquier caso–, un borrador sin pulir de los hallazgos importantes, y después decidir si presentar los datos en secciones separadas o divididos para poder integrarlos con otros datos. (Para obtener mayor información al respecto, véase el módulo 32.)

V. ESTRATEGIAS PARA COMPROBAR O CONFIRMAR LOS HALLAZGOS Y PROBAR SU VALIDEZ

Los investigadores que utilizan diseños de investigación cuantitativos reducen los datos a cifras y aplican pruebas estadísticas, lo que no garantiza la validez de sus resultados: algo puede haber salido mal durante el muestreo o la recolección de los datos, o incluso en el primer diseño del estudio (en que se pudo haber pasado por alto alguna variable de confusión). Las siguientes estrategias pueden ser, por consiguiente, útiles para todo investigador, pero son particularmente importantes en la investigación cualitativa, puesto que bastan los números pequeños de la investigación cualitativa generan frecuentemente dudas respecto de su validez.

1. *Compruebe* la representatividad de los datos. Aunque en la investigación cualitativa no suele seleccionarse a los informantes de manera aleatoria, la selección debe haberse hecho sistemáticamente, de acuerdo con reglas establecidas de antemano (véase el módulo 11). Compruebe que haya entrevistado efectivamente a todas las categorías de informantes necesarias para completar un panorama íntegro de su tema (y que no dependa demasiado de las autoridades comunicativas). Cerciórese de no generalizar a partir de acontecimientos que no sean realmente representativos.
2. *Revise que no haya sesgos* debido a sesgos del observador o a la influencia del investigador en la situación de investigación. Ya comentamos los pormenores al respecto en el *módulo 10*.
3. *Verifique los datos con evidencia de otras fuentes independientes*. Estas otras fuentes pueden ser distintos informantes independientes; el empleo de diferentes técnicas de investigación para el mismo tema, o los resultados de otros estudios similares (véanse los módulos 5, 9 y 10.) Los datos deben confirmarse o por lo menos no contradecirse entre sí.

La verificación activa de los datos de la investigación con evidencia independiente o contra-evidencia es una de las mejores maneras de asegurar su validez.

Por ejemplo, las respuestas de uno y otro cónyuges (y otros informantes involucrados) sobre temas tales como quién decide si emplear métodos anticonceptivos y cuáles deberían emplearse, sobre quién decide si los hijos deben ser circuncidados, o sobre qué ha cambiado en la relación conyugal tras el diagnóstico de la lepra o alguna otra enfermedad temible en uno de los cónyuges, deben confirmarse entre ellos.

4. *Compare y contraste los datos*.

Muchas veces la comparación está implícita en el diseño de la investigación mediante la inclusión de diferentes categorías de informantes.

Si queremos estar seguros, *por ejemplo*, de que la variable A (alto nivel de educación) influye en la variable B (uso de métodos de planificación familiar), tenemos que comparar a dos grupos de madres –uno con educación superior y otro con un bajo nivel educativo–, sobre su uso de métodos de planificación familiar.

Es importante que compare y contraste los datos si desea identificar sus variables y *confirmar* las asociaciones entre ellas.

5. *Use al máximo (grupos de) informantes situados en los extremos*

En la discusión del diseño del estudio y el muestro determinamos que puede ser útil buscar categorías de informantes que representen los extremos de una determinada variable.

Por ejemplo, puede usted descubrir que es más útil estudiar a los “desertores” y a los asistentes regulares a los servicios de tuberculosis, y dejar de lado la categoría de los asistentes irregulares. Ésta puede ser la forma más eficaz de identificar las principales variables que influyen en el comportamiento de los pacientes con tuberculosis respecto de su asistencia.

6. *Haga una investigación adicional para evaluar los hallazgos de su estudio.*

Los resultados de su estudio pueden ser tan desconcertantes que usted decida hacer después un estudio de seguimiento. Dicho estudio puede realizarse por varias razones:

- para reproducir algunos de los hallazgos
- para descartar (o identificar) las posibles variables de intervención
- para descartar explicaciones antagónicas mediante la investigación
- para buscar evidencia negativa

Los estudios adicionales que realice por una o varias de estas razones pueden servirle para hacer más convincentes los resultados de su estudio inicial.

7. *Obtenga comentarios de sus informantes.*

A lo largo de los *módulos 1-20* hemos insistido en que es necesario involucrar a todos los sectores interesados en las diferentes etapas de la investigación. Esto es importante no sólo por razones éticas o porque hacerlo aumenta las probabilidades de aplicar los resultados, sino también porque *mejora la calidad del diseño de su estudio, de sus datos y de las conclusiones derivadas de dichos datos*. Las sugerencias y la información adicional recogida en las sesiones de discusión van a mejorar invariablemente la calidad de su informe de investigación.

TRABAJO EN EQUIPO (El tiempo necesario depende de la cantidad de los datos cualitativos recogidos.)

1. *Revise si enlistó todas las fuentes de datos cualitativos* para cada objetivo cuando hizo un inventario de sus datos en la sesión de trabajo en equipo del *módulo 21*.
2. *Cerciórese de haber terminado la categorización de las respuestas a todas sus preguntas abiertas* y de haberlas incluido en sus hojas maestras o en su análisis en la computadora, junto con otras series de datos. Analice el contenido de las respuestas e incluya los datos relevantes en las recomendaciones para la acción o los planes de acción subsecuentes (vea el *módulo 33*).

3. *Describa sus muestras* si su estudio incluyó GD, entrevistas con informantes clave u observaciones.
4. Organice estos datos por tema; de ser necesario codifíquelos y *transcríbalos por tema en hojas de compilación*.
5. *Decida* si va a emplear matrices, diagramas u hojas de rota-folios para sintetizar sus datos.
6. *Interprete los datos* comparando a los diferentes grupos de participantes en los GD o a los informantes clave (si los tiene) y vea de qué manera responden a sus objetivos de investigación.
7. *Enliste los principales hallazgos y conclusiones derivados de los datos cualitativos* y determine de qué manera complementan los datos de otras fuentes consideradas en su estudio.
8. *Decida cómo incorporar los datos en su informe*: ya sea en una sección o integrándolos a los hallazgos recabados mediante otras técnicas de recolección de datos. Decida qué debe entrar en la discusión y qué material es útil para desarrollar intervenciones.
9. *Verifique sus conclusiones* (vea la sección V) y decida si hay algunas que todavía quisiera comprobar y cómo hacerlo.

Referencias

- Miles MB and Huberman AM (1984). *Qualitative data analysis, a sourcebook of new methods*, Beverley Hills, CA, USA.: Sage Publications.
- Patton MQ (1990) *Qualitative Evaluation and Research Methods*, 2nd ed., Newbury Park, CA: Sage Publications.
- Spradly JP (1979). *The ethnographic interview*, New York, NY, USA: Holt, Rinehart and Winston.
- Walker R (ed) (1985). *Applied qualitative research*, Hants, UK: Gower Publishing Company Ltd.
- Willms DG and Johnson NA (1996). *Essentials in Qualitative Research: A Notebook for the Field*. Hamilton, Canada: Mc Master University.
- Yin RK (1984). *Case study research: design and methods*. Beverly Hills, CA, USA.: Sage Publications.

NB: Una fuente de inspiración muy importante para la escritura de este módulo fue el libro de Miles y Huberman. La sección V de este módulo es una versión muy abreviada y adaptada de su capítulo VII.

Anexo 23.1 Ejemplo de una hoja de compilación (sexo y lepra)

No.	Sexo	Edad	Educ.	Casado/a	Estatus econ.	Síntomas		Primeras reacciones (10)	Estigma sufrido en la realidad				Ec. doméstica (12)	Percepción de la cura (25)
						Cuando el diagnóstico (6)	Ahora (18)		Cónyuge	Parientes	Parientes políticos	Comunidad		
1	M	40	6 años	Sí	Agricultor Tendero	-Parches -Dolor en los nervios -Pérdida del pie (2)	Todavía ninguno	Mucho temor -La esposa va a huir -La comunidad lo va a aislar -Se le van a caer los dedos y los pies -Ya no va a poder trabajar y mantener a fam.	-Mantiene su apoyo -Ayuda más en la compra -decidió abstenerse en lo sexual. (8 meses) Autoestigma	-Los niños apoyan; los pequeños no se dan cuenta Padres y hns. Visitan + que antes	No lo dijo Lo oculta	-No lo dijo -Cree que los amigos no lo saben -Se comportan como antes Lo oculta ¿No hay estigma?	Tiene ayuda doméstica (No está obligado/a a trabajar) Ingreso	No (todavía hay signos: caída del pie)
10	F	21	8 años	Sí	Esp. agric Padre gran agric	Parches (la maestra lo vio y la refirió al CS)	Ninguno	Sabía el título; preocupado -Mala enfermedad -Novio va a cancelar la boda	-Novio idaga en el CS -si es ucrable -si podrá procerar -Boda pospuesta -Ahora tiene hijos Estigma retraído	-Los padres apoyan mucho	Primera esposa lo dijo a todo el mundo	Todos los del lugar asistieron a la boda. No hay estigma	Lo hace todo	Sí (no hay signos)
12	F	60	...	Sí Pero ahora divorciada	Pequeña agric y comercio	Parches		Preocupada -Mala enfermedad -Conyuge enojado Síntomas	-Conyuge la echó. Divorcio	El hijo la llevó a su casa. Apoyo	Suegro le da su apoyo	Evitó las grandes reuniones pero ahora bien. El autoestigma se retajo	Comercio a pequeña escala para ganar el importe del autobús para ir a tratamiento	Sí, curado/a pero cony.murió

Módulo 23 ANÁLISIS DE LOS DATOS CUALITATIVOS

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
1 hora	Trabajo en equipo

(El tiempo debe ajustarse dependiendo de la cantidad y el tipo de los datos cualitativos.)

Introducción y discusión

- Si ninguno de los grupos tiene más datos cualitativos que unas cuantas preguntas abiertas de los cuestionarios, quizás prefiera usted concentrarse en la sección II para darles a los participantes un panorama general de cómo se podrían procesar los datos cualitativos, y dar sólo una breve pasada por las secciones III y IV.
- Pero si los participantes del curso tienen experiencia/ capacitación/ interés en la investigación, podría usted cubrir completas las secciones III y IV, aun cuando no hayan recogido una gran cantidad de datos cualitativos. Los procedimientos presentados para esbozar las conclusiones y comprobar su validez son pertinentes en cualquier tipo de investigación, y los métodos para comprobar y contrastar los datos pueden no ser del conocimiento de todos los participantes.
- Haga referencia al análisis del diagrama de flujo que hizo cada grupo cuando prepararon sus propuestas de investigación, al que preceden los módulos 1-18 y 22-33, y a cualquier otro ejemplo de diagrama o gráfico que considere ilustrativo en láminas de retroproyector o en hojas de rota-folios.
- Si uno o varios grupos hicieron una extensa investigación cualitativa, cubra meticulosamente el módulo, utilizando ejemplos de sus estudios. Muy probablemente ninguno de los participantes va a estar muy familiarizado con el análisis de los datos cualitativos.
- Permita que los grupos que hicieron una investigación cualitativa describan en la sesión plenaria cómo analizaron los datos de los grupos de discusión, de las observaciones, o de las entrevistas con los informantes clave. Pídales que describan qué preguntas agregaron, o cuáles descartaron en el transcurso de las sucesivas entrevistas, y por qué.

Trabajo en equipo

- *Para todos los grupos:*

Revise si cualquiera de sus preguntas abiertas necesita un análisis del contenido de las respuestas individuales. Algunas preguntas de opinión podrían aportar un valioso material ilustrativo para sus informes. Tome nota de esto, puesto que los grupos podrían olvidar dichos datos cuando entren de lleno en las tablas y las estadísticas. Comente si basta con enlistar los planteamientos para analizar el contenido, o si sería deseable exponer gráficamente los datos.

En cuanto a los grupos que tengan datos cualitativos elaborados a partir de los GD o de las entrevistas a los informantes clave: revise con ellos todos los datos disponibles y asesórelos para que sigan las direcciones del trabajo en equipo.

Este espacio en blanco intencionalmente dejado de la página

Módulo 24

TABULACIÓN CRUZADA DE LOS DATOS CUANTITATIVOS

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 24 TABULACIÓN CRUZADA DE LOS DATOS CUANTITATIVOS

OBJETIVOS

Al terminar esta sesión, usted debe ser capaz de:

1. *Describir* la diferencia entre la tabulación cruzada descriptiva y la analítica.
2. *Construir* todas las tabulaciones cruzadas importantes que contribuyan a alcanzar sus objetivos de investigación.
3. *Interpretar* las tabulaciones cruzadas en relación con sus objetivos y las preguntas de su estudio.

- I. Introducción.
- II. Los diferentes tipos de tabulaciones cruzadas.
- III. Construcción de tabulaciones cruzadas adecuadas para los objetivos de investigación.
- IV. Pistas generales para construir tablas.
- V. Interpretación de los resultados.

I. INTRODUCCIÓN

Hasta aquí hemos hecho tablas que contienen distribuciones de frecuencias para *una variable a la vez*, a fin de describir parcialmente nuestros datos. Dependiendo de los *objetivos* de nuestro estudio y del *tipo* de estudio, podríamos tener que *revisar las relaciones entre varias de nuestras variables* a la vez, para poder describir adecuadamente nuestro problema o identificar sus posibles explicaciones.

Para este propósito es adecuado construir TABULACIONES CRUZADAS.

II. LOS DIFERENTES TIPOS DE TABULACIONES CRUZADAS

Dependiendo de los objetivos y el tipo de estudio, pueden ser necesarias diferentes clases de tabulaciones cruzadas:

- Hay tabulaciones cruzadas cuyo propósito es *describir el problema de estudio* presentando una combinación de variables. Los estudios exploratorios y descriptivos deben generar dichas tablas, pero también en los estudios analíticos es necesario primero describir el problema (véanse las tablas 24.2 y 24.3).
- Otras tabulaciones cruzadas *analizan las relaciones* entre las variables independientes y las dependientes, *es decir, entre los problemas definidos y los factores que contribuyen a éstos*. Los estudios exploratorios que identifican dichos factores pueden generar tablas cruzadas analíticas; los estudios comparativos transversales, los de casos y controles y los de cohorte las necesitan de hecho (véase la identificación de asociaciones en la tabla 24.6 y las tablas 24.4 y 24.5).
- Están también las tablas cruzadas, cuyo objetivo es *comparar el resultado de las intervenciones o experimentos* en un grupo de estudio que haya participado en la intervención, y un grupo de control que no haya participado. Así se determina si dicha intervención estableció una diferencia. Éste es el procedimiento estándar en los estudios experimentales y los cuasi-experimentales (véase la tabla 24.7).
- Unas tablas cruzadas muy comunes, necesarias en *cualquier* estudio, son las que *describen las muestras* en las que se concentra la investigación, tomadas de las poblaciones estudio. Con éstas vamos a comenzar.

1. Tabulaciones cruzadas para describir la muestra

En cualquier estudio, a pequeña escala (exploratorio) o a gran escala, es común describir a los sujetos de investigación incluidos en las muestras antes de presentar los resultados reales del estudio. Esto puede hacerse para variables por separado, en una simple tabla de frecuencia (como se muestra en el *módulo 22*) o para una combinación de variables, en una tabla cruzada.

Ejemplo 1:

Se llevó a cabo un estudio sobre el grado de satisfacción laboral entre los médicos y las enfermeras en zonas rurales y urbanas. Para describir la muestra se construye una tabulación cruzada con el sexo y la residencia (rural o urbana) de los/as médicos/as y los/as enfermeros/as entrevistados. Esto es útil porque las opiniones del personal masculino y femenino se comparan por separado para las zonas rurales y las urbanas.

Tabla 24.1a Tipo de trabajador de la salud por lugar de residencia

Residencia	Tipo de trabajador de la salud		Total
	Médicos/as	Enfermeros/as	
Rural	10 (16%)	69 (38%)	79 (33%)
Urbano	51 (84%)	113 (62%)	164 (67%)
Total	61 (100%)	182 (100%)	243 (100%)

La tabla 24.1a muestra que un porcentaje más alto de enfermeros/as que de médicos/as trabajan en zonas rurales, pero que, en general, una mayor proporción del personal trabaja en zonas urbanas (67%).

Tabla 24.1b Sexo de los trabajadores de la salud por lugar de residencia

Residencia	Sexo del trabajador de la salud		Total
	Hombres	Mujeres	
Rural	54 (43%)	25 (21%)	79 (33%)
Urbano	71 (57%)	93 (79%)	164 (67%)
Total	125 (100%)	118 (100%)	243 (100%)

A partir de la *tabla 24.1b* puede concluirse que hay más hombres que mujeres que prestan sus servicios en las zonas rurales. Estos hombres en las zonas rurales son aparentemente enfermeros.

Para obtener un panorama general de la distribución por sexo de los médicos y las enfermeras en las zonas rurales y las urbanas, podemos construir las siguientes tablas cruzadas de dos por cuatro.

Tabla 24.1c Residencia y sexo de médicos/as y enfermeros/as

Trabajadores de la salud		Residencia		Total	
		Rural	Urbana		
Médicos/as	Hombres	8 (10%)	35 (21%)	43 (18%)	
	Mujeres	2 (3%)	16 (10%)	18 (7%)	
Enfermeros/as	Hombres	46 (58%)	36 (22%)	82 (34%)	
	Mujeres	23 (29%)	77 (47%)	100 (41%)	
Total		79 (100%)	164 (100%)	243 (100%)	

Esta tabla nos muestra en forma rápida que, en efecto, los enfermeros dominan en los servicios de salud rurales. También indica que los hombres dominan en la profesión médica, (18% médicos<-> 7% médicas), pero que hay en general más enfermeras que enfermeros, y que las enfermeras constituyen principalmente el grupo urbano.

Los datos presentados en tablas se enlistan por lo general en cifras absolutas así como en frecuencias relativas (porcentajes o proporciones).

Como ya se vio en el *módulo 22*, la media, la mediana o el modo de los datos numéricos (tales como la edad), se calculan también para describir la muestra.

2. Tablas descriptivas de un problema

Las tablas cruzadas son muy útiles para describir el problema de estudio.

Ejemplo 2:

Queremos saber la edad en que ocurren los *embarazos de adolescentes*, y si son más frecuentes entre las estudiantes que entre las jóvenes que no asisten a la escuela. Para responder a estas preguntas, construimos la siguiente tabulación cruzada. (Los datos son imaginarios.)

Tabla 24.2 Número de embarazos a diferentes edades entre jóvenes que asisten a la escuela y jóvenes que no lo hacen (Provincia X, 1998-2000)

Edad en la concepción	Número de embarazos		Total
	Jóvenes que asisten a la escuela N = 500	Jóvenes que no asisten a la escuela N = 500	
12 años	2 (3%)	1 (1%)	3 (2%)
13 años	2 (3%)	3 (3%)	5 (3%)
14 años	5 (7%)	12 (12%)	17 (10%)
15 años	23 (34%)	35 (35%)	58 (10%)
15 años	36 (53%)	48 (49%)	84 (50%)
Total	68 (100%)	99 (100%)	167 (100%)
Prevalencia	68/500 13,6%	99/500 19,8%	167/1000 16,7%

La *tabla 24.2* revela que los embarazos se producen de los 12 años en adelante, pero que de los 15 años en adelante el problema se agudiza considerablemente tanto en las jóvenes que asisten como en las que no asisten a la escuela. Sin embargo, el porcentaje de embarazos es mayor entre las jóvenes que no van a la escuela (casi 20%) que entre las que van a la escuela (13,6%), y el primer grupo parece embarazarse a una edad ligeramente más temprana.

Ejemplo 3:

Se hizo un estudio para examinar los factores que contribuyen a la alta proporción de alumbramientos de mortinatos en un hospital. La siguiente tabulación cruzada describe cuántos de los mortinatos frescos y macerados (perdidos) pesaron menos de 2500 gramos y cuántos 2500 gramos o más.

Tabla 24.3 **Peso del feto por su condición al nacer**

Peso del feto	Condición al nacer		Total
	Frescos	Macerados	
Menos de 2500 gramos	9 (10%)	20 (38%)	29
2500 gramos o más	79 (90%)	12 (62%)	91
Total	88 (100%)	32 (100%)	120

En la *tabla 24.3* puede verse que la mayoría de los niños nacidos muertos frescos (90%) eran de peso normal, mientras que una gran proporción (63%) de los mortinatos macerados tenía un peso bajo al nacer.

Note que en las tabulaciones cruzadas descriptivas, es conveniente poner en las columnas a los grupos que se describen (condición al nacer, en la *tabla 24.3*)

En las *encuestas* de corte transversal, es posible jugar con las variables e identificar posibles asociaciones o interrelaciones.

Ejemplo 4:

En una *encuesta de corte transversal sobre desnutrición*, por ejemplo, podrían ponerse a prueba las relaciones entre la duración de la lactancia materna y la edad de la madre, o su estatus laboral (en respuesta a preguntas de investigación previamente formuladas, pero a veces a nuevas preguntas surgidas durante el análisis del material).

Tome en cuenta que estas tablas le permiten calcular los porcentajes tanto horizontal como verticalmente, puesto que todas las variables tienen una oportunidad similar de aparecer en la encuesta. Sin embargo, por lo general ponemos en las filas la variable de la que suponemos que influye en la otra, mientras que la variable “dependiente” va en las columnas (*véanse* las *tablas 24.4 y 24.5*).

La *tabla 24.4* indica que las madres más jóvenes amamantan durante un periodo más corto que las de más edad, puesto que el porcentaje más alto de madres que amamantan sólo 0-5 meses es el grupo de edad más joven (15-19). Con la edad disminuye el porcentaje de las madres que destetan a sus hijos antes de los 6 meses. Para las madres que amamantan durante más de 12 meses se observa exactamente la tendencia contraria: el porcentaje más alto (70%) está en

el grupo de las madres de más edad (mayores de 40 años), mientras que sólo el 10% de las madres de 15-19 años amamantan durante más de 12 meses. Así, parece haber una asociación entre la edad de la madre y la duración de la lactancia materna.

Tabla 24.4 Duración de la lactancia materna por edad de la madre

Edad (en años)	Duración de la lactancia materna			Total
	0 - 5 meses	6 - 11 meses	12 > meses	
15 -19	18 (62%)	8 (28%)	3 (11%)	29 (100%)
20 -24	27 (44%)	25 (40%)	10 (16%)	62 (100%)
25 -29	15 (18%)	33 (40%)	35 (42%)	83 (100%)
30 - 34	5 (14%)	13 (37%)	17 (49%)	35 (100%)
35 - 39	2 (10%)	7 (33%)	12 (57%)	21 (100%)
40 +	0 (0%)	3 (30%)	7 (70%)	10 (100%)
Total	67 (28%)	89 (37%)	84 (35%)	240 (100%)

Si desea determinar si hay una asociación entre el estatus laboral de las madres y la duración de la lactancia materna, la *tabla 24.5* es la apropiada.

Tabla 24.5 Estatus laboral de las madres en relación con la duración de la lactancia

Estatus laboral de la madre	Duración de la lactancia materna			Total
	0 - 5 meses	6 - 11 meses	12 + meses	
Tiempo completo	56 (42%)	49 (37%)	27 (49%)	132 (100%)
Tiempo parcial	5 (21%)	15 (33%)	4 (57%)	24 (100%)
No trabaja	6 (7%)	25 (30%)	53 (70%)	84 (100%)
Total	67 (28%)	89 (37%)	84 (35%)	240 (100%)

Las mujeres que trabajan tiempo completo amamantan durante un periodo más corto que las madres que no trabajan. Por lo tanto, parece haber una asociación entre el estatus laboral de las madres y la duración de la lactancia.

Aquí nos encontramos en un área gris entre las tablas descriptivas y las analíticas, pero ya vimos en el *módulo 9* (tipos de estudio) que las encuestas de corte transversal pueden convertirse fácilmente en estudios comparados (analíticos) si se recogen los datos suficientes sobre los factores que podrían contribuir al problema en estudio (véase más abajo).

3. Tabulaciones cruzadas analíticas

En los estudios *comparativos transversales* y *de casos y controles* comparamos dos grupos, uno con un problema dado y otro sin él, para identificar las variables independientes que contribuyen al problema.

Recordará usted (del *módulo 9*) que la comparación transversal se basa en una encuesta de corte transversal mayor (que es un estudio descriptivo). Sin embargo, puede constituir la base de un estudio comparado al seleccionar los grupos extremos de la muestra de la encuesta, uno con un problema específico (como la desnutrición grave) y un grupo de control (de niños bien alimentados) para identificar los factores que contribuyen al problema (de desnutrición).

Ejemplo 4 (continuación)

Uno de los posibles factores que contribuyen a la desnutrición de los niños menores de 5 años es el desconocimiento de las madres acerca de los alimentos apropiados para el periodo de destete. Un estudio comparativo transversal sobre la desnutrición, basado en la encuesta, arrojó los siguientes resultados:

Tabla 24.6 Nivel de conocimiento de las madres y estatus nutricional de sus hijos

Nivel de conocimiento nutricional de la madre (alimentos para el destete)	Estatus nutricional de los niños		Total
	Gravemente desnutridos (casos)	Bien nutridos (controles)	
Bajo	45 (69%)	10 (15%)	55
Alto	20 (31%)	55 (85%)	75
Total	65 (100%)	65 (100%)	130

Parece que las madres de los niños gravemente desnutridos tienen un menor conocimiento de los alimentos para el periodo de destete que las madres de los niños bien nutridos.

Pueden construirse otras tabulaciones cruzadas analíticas para el estudio mencionado en el *ejemplo 4*. Los dos grupos (niños gravemente desnutridos y niños bien nutridos) pueden disponerse sistemáticamente en las columnas. Y en las filas pueden ponerse entonces diferentes variables independientes, tales como la fuente de agua potable (protegida o sin protección) o el estatus de inmunización (con la vacunación completa o no).

Los *estudios de casos y controles* siguen los mismos procedimientos de la construcción de tabulaciones cruzadas, como se muestra en la *tabla 24.6*.

En los *estudios de cohorte* las tabulaciones cruzadas se construyen igual que en los estudios de control de caso o en los comparativos transversales. Los factores de riesgo previamente definidos se colocan en las filas, mientras que los participantes en el estudio que contraen determinadas afecciones o enfermedades y los que no se ven afectados se colocan en las columnas de la tabla.

En los *estudios cuasi-experimentales o experimentales*, el investigador compara dos grupos similares, (uno de ellos sometido a una intervención y el otro no) antes y después de la intervención, para medir el efecto de ésta.

Si el estudio comparativo transversal mencionado en el *ejemplo 4* da lugar a una intervención de capacitación para la salud, sería conveniente llevar a cabo un estudio cuasi-experimental mediante el cual se pueda medir el efecto de dicha capacitación (véase la tabla 24.7).

Tabla 24.7 Asistencia de las mujeres a la educación nutricional y su nivel de conocimiento de nutrición

Asistencia a la capacitación nutricional	Nivel del conocimiento en nutrición						Total
	Bajo		Promedio		Alto		
Asistentes	20	(13%)	90	(60%)	40	(27%)	150 (100%)
No-asistentes	40	(45%)	30	(33%)	20	(22%)	90 (100%)
Total	60	(25%)	120	(50%)	60	(25%)	240 (100%)

La *tabla 24.7* indica que las mujeres que asistieron a las charlas de nutrición tienen un mayor nivel de conocimientos del tema que las que no asistieron. Parece, por consiguiente, que la intervención fue efectiva.

Tome en cuenta que en la tabulación cruzada anterior, los grupos que debían compararse se pusieron en las filas, mientras que los diferentes niveles de conocimiento nutricional se pusieron en las columnas. Esto se debe a que el conocimiento nutricional es resultado de la asistencia a las charlas de nutrición y se considera, por consiguiente, como la variable dependiente. Los porcentajes se calculan horizontalmente, pues el interés está en comparar el nivel de conocimiento entre quienes asistieron y quienes no asistieron.

Para que sea más fácil interpretar los datos y redactar el informe de una manera sistemática, es recomendable colocar las variables dependientes en las columnas y las variables independientes en las filas.

Observe la diferencia entre las *tablas 24.6 y 24.7*. En la *tabla 24.6* el “conocimiento en nutrición” es la *variable independiente* y, por lo tanto, se coloca en las filas, y los dos grupos que se comparan se colocan en las columnas. En cambio, en la *tabla 24.7* el “conocimiento en nutrición” se convirtió en la *variable dependiente*, influida por la intervención y, por lo tanto, aparece en las columnas.

Al calcular los porcentajes para la tabla cruzada, es importante recordar que el total de cada uno de los grupos que se van a comparar debe ser 100%.

III. CONSTRUCCIÓN DE TABULACIONES CRUZADAS ADECUADAS PARA LOS OBJETIVOS DE INVESTIGACIÓN

Cuando diseñó su proyecto de investigación proyectó un plan de cruces con sus tablas correspondientes para los datos que pretendía recoger (*módulo 13*) sobre la base de los objetivos y el tipo de estudio.

Ahora que ya recogió los datos y tiene una idea de su calidad y de cómo pueden utilizarse, examine de nuevo, sistemáticamente, las tabulaciones cruzadas por realizar.

Para construir tabulaciones cruzadas adecuadas, le recomendamos seguir los pasos que se describen a continuación:

1. Revise cada objetivo específico y el método de recolección de los datos correspondientes.
2. Formule frases hipotéticas de las conclusiones a las que espera llegar con respecto a cada objetivo.

Por ejemplo, en la encuesta de corte transversal sobre la desnutrición de niños menores de 5 años, las conclusiones hipotéticas para el objetivo específico de determinar los factores asociados con el destete precoz, pueden ser:

- Las madres que trabajan destetan a sus hijos antes que las madres que no tienen empleo
- Las madres que no asisten a charlas de nutrición destetan a sus hijos antes que las madres que sí asisten

Formular conclusiones posibles lo ayudan a:

- recordar el propósito de cada tabulación y cada cálculo que haga
 - evitar perder tiempo en cálculos y tabulaciones poco significativos
 - mantener sus datos organizados de modo que le sea más fácil escribir un informe bien organizado
3. Para cada conclusión hipotética, construya una tabla cruzada hipotética para derivar las conclusiones correctas.
 4. Haga el recuento adecuado de la frecuencia (utilizando las hojas maestras) e ingrese los resultados en las celdas.
 5. Interprete los datos y escriba conclusiones claras. No es necesario ni recomendable describir el contenido de cada celda de la tabla ya que esto aburre a los lectores. (Vea la interpretación de las tablas presentada anteriormente en este módulo).

EJERCICIO 1:

Seleccione un objetivo específico de cada proyecto de investigación, formule las conclusiones esperadas y construya las eventuales tablas cruzadas.

IV. PISTAS GENERALES PARA CONSTRUIR TABLAS

- Cerciórese de que todas las categorías de las variables presentadas en las tablas se hayan especificado y sean exhaustivas y mutuamente excluyentes (es decir, sin superposiciones ni vacíos).
- Compruebe que el recuento de las columnas y las filas corresponda a los cálculos de la frecuencia de cada variable.
- Compruebe también que el gran total de la tabla corresponda al número de personas de la muestra. De no ser así, es necesaria una explicación. Ésta podría presentarse como una nota al pie de página. (Datos perdidos, por ejemplo).
- Piense en un título claro para cada tabla. Cerciórese también de que los encabezados de las filas y las columnas no den lugar a interpretaciones erróneas.
- Numere sus tablas y téngalas junto a los objetivos con los que se relacionan. Esto lo va a ayudar a organizar su informe y a asegurar que el trabajo no se duplique.

V. INTERPRETACIÓN DE LOS RESULTADOS

Todas las tablas presentadas en este módulo van acompañadas de una interpretación. Usted habrá notado que *describir a los informantes* en términos de profesión, ubicación, edad o sexo es muy sencillo. Habrá notado también, diferencias entre los grupos de informantes en función de estas variables contextuales, interesantes como para continuar *explorarándolas*. Por ejemplo, si entre los niños desnutridos predominan los de cierta zona, o los de sexo femenino, usted puede explorar más a fondo las razones tras esta diferencia.

En cuanto a las *asociaciones entre las variables*, es difícil identificar *en qué dirección van las asociaciones y lo que esto significa para el problema de estudio*. Si hay una asociación entre el estatus laboral de una madre y sus prácticas de destete, puesto que el destete precoz (= alimentación de biberón prematura) se asocia a la desnutrición, especialmente entre las madres de escasa educación, podría ver más a fondo a las madres que trabajan tiempo completo. ¿Cuál es su nivel educativo? Parecería que las madres sin educación o con escasa educación, que trabajan tiempo completo podrían constituir un grupo de riesgo específico para la desnutrición. Esto debe revisarse. Este nivel de atención en los datos es lo que convierte al investigador en un “detective” (véase el módulo 10B parte V, habilidades para la entrevista).

La interpretación de los datos debe *formularse correctamente*. Debe prestarse particular atención a la redacción de las conclusiones derivadas de las tablas en

que se *comparan determinados grupos seleccionados* con respecto a una serie de variables independientes. En este caso, exprese la incidencia de las variables independientes dentro de los grupos seleccionados, y no al revés.

Por ejemplo, en la *tabla 24.6* usted va a afirmar que las madres de los niños desnutridos parecen saber menos de los alimentos apropiados para el periodo de destete que las madres de los niños bien nutridos. Usted *no* va a afirmar que entre las madres con un bajo nivel de conocimiento hay un mayor porcentaje (82%) de niños desnutridos que de niños bien nutridos. Esto se debe a que usted seleccionó en su encuesta transversal poblaciones a las que quiere comparar sobre la base del *estatus nutricional*. De ahí que pueda haber distorsionado la distribución de las madres con un nivel alto o un nivel bajo de conocimiento de los alimentos apropiados para el periodo de destete (esas madres no tienen la misma oportunidad de aparecer en su sub-muestra).

Para ayudarle a formular sus resultados de manera correcta y consistente, se ha subrayado que cada uno de los grupos que compare debe sumar un total de 100% con respecto a las variables con que los compare.

Necesidad de pruebas estadísticas adicionales

Habrán notado que todas las interpretaciones de las tablas se han expresado en términos provisionales: “*parece que...*” Aun cuando *parece* haber una relación entre las variables, o una gran diferencia entre los grupos comparados –en la ausencia o en la presencia de un determinado factor–, podemos estar seguros sólo si *una prueba estadística confirma* esta relación. Los *módulos 28-31* muestran los procedimientos estadísticos necesarios para ello.

Variables de confusión

En los *módulos 8 y 9* comentamos que a veces puede parecer que hay una relación entre dos variables, pero que en realidad esta relación se ve perturbada por otra variable de confusión que se relaciona con ambas.

Esto es particularmente molesto cuando intentamos resolver un problema y buscamos identificar los factores que contribuyen a producirlo. Las variables de confusión nos pueden llevar a esbozar conclusiones erróneas y, por consiguiente, a tomar decisiones equivocadas para la acción.

Así, en el ejemplo dado en el *módulo 8*, se menciona *la educación* como una posible variable de confusión que desdibuja la relación entre las prácticas de las madres en la alimentación con biberón, y la prevalencia de diarrea en los niños. En el *módulo 26* se demuestra que si se considera tan sólo la relación entre el inicio de la alimentación con biberón y la diarrea de los niños, no se encuentra asociación alguna: la alimentación prematura con biberón parece incluso reducir

la incidencia de la diarrea. Sin embargo, cuando el grupo de las madres se estratificó (se dividió) en dos grupos según su nivel de educación, sí pareció haber una relación. En el grupo de las madres de escasa educación que comenzaron pronto la alimentación con biberón hubo una incidencia de diarrea en los niños menores de dos años significativamente mayor que en el grupo de las madres del mismo nivel de educación que destetaron a sus hijos más tarde. Entre las madres con una educación superior, el introducir la alimentación con biberón no dio lugar a ninguna diferencia: el número de episodios de diarrea en los niños fue bajo de principio a fin, porque al parecer siguieron mejores prácticas de higiene. *La educación era, por lo tanto, una variable de confusión.* El problema entre las madres de escasa educación no se habría hecho visible si no hubiéramos estratificado el grupo de las madres según el nivel educativo.

Al interpretar las tabulaciones cruzadas, debemos estar siempre alertas ante las posibles variables de confusión. En el *módulo 26* comentamos los procedimientos para controlar las variables de confusión. También vamos a comentar el análisis de los *datos apareados*, recogidos para evitar el efecto de confusión a partir de variables específicas como la educación o la edad.

TRABAJO EN EQUIPO

- Revise cada objetivo específico y su correspondiente diseño de investigación: formule oraciones hipotéticas que describan el tipo de conclusiones a las que espera llegar respecto de cada objetivo.
- Construya tabulaciones cruzadas hipotéticas considerando si desea:
 - describir a los sujetos de investigación de su muestra o describir el problema
 - comparar a los grupos para encontrar diferencias
 - encontrar asociaciones entre las variables

Remítase a las tabulaciones cruzadas hipotéticas que hizo en el primer taller (módulo 13).

- Si construye tabulaciones cruzadas analíticas, trate de identificar las posibles variables de confusión.
- Por último, llene las tabulaciones cruzadas hipotéticas con los datos, calcule los porcentajes e interprete lo que pueden significar en relación con sus objetivos y sus preguntas de estudio.

Módulo 24 TABULACIÓN CRUZADA DE LOS DATOS

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
3 horas +	Trabajo en equipo
1 hora	Presentaciones del grupo y discusión plenaria (opcional)

Introducción y discusión

- Se recomienda que utilice un retroproyector de transparencias (o un rotafolio) para captar la atención de los participantes cuando presente y explique la construcción de las diferentes tablas. No se limite a remitirse a los módulos.
- Algunas de las tabulaciones cruzadas presentadas en el módulo se llenaron con datos imaginarios. Esto se hizo para tener ejemplos más concretos y mostrar cómo debían interpretarse las tablas. Se debe prestar especial atención al diseño de las tablas (qué va en las filas y qué en las columnas). Por lo tanto, se recomienda que use dos transparencias, que puede colocar una encima de la otra, para presentar la tabla: una con las tablas hipotéticas, y la otra con los datos propiamente dichos.
- Preste mayor atención a la manera en que debe leerse cada tabla: algunas deben leerse horizontalmente y otras verticalmente, dependiendo de si los grupos que se están comparando se pusieron en las filas o en las columnas.
- El ejercicio hecho en clase al final de la sección III ayuda a darles a los participantes una práctica útil para diseñar tablas cruzadas adecuadas. Trate de obtener al menos una tabla de cada uno de los tres tipos (el tipo de tabla descriptiva y los dos tipos de tabla analítica) de los grupos en función de sus estudios.
- Cerciórese de que todos estén de acuerdo en cómo leer la tabla (horizontal o verticalmente). Si los grupos tienen dificultades para construir tablas adecuadas para sus propios proyectos, esto indica que les conviene presentar algunas de sus tabulaciones cruzadas en la sesión plenaria, después de la sesión de trabajo en equipo, para recibir asesoría.
- Las pistas para construir tablas (sección IV del módulo) deberían ilustrarse con un ejemplo de una tabla, tomado del módulo o de alguno de los grupos.

Trabajo en equipo

- Cubra los tres primeros pasos de la asignación del trabajo en equipo con el grupo como un todo. Pueden formarse subgrupos para llenar las tablas con los datos correspondientes.
- Subraye que una vez llenadas las tablas con los datos, pueden interpretarse de inmediato. Las conclusiones preliminares deben registrarse debajo de cada tabla para facilitar la redacción del informe que se hará durante la segunda semana. Estas conclusiones deben comentarse más adelante con los demás miembros del grupo.
- En esta etapa la numeración de las tablas puede hacerse de acuerdo con el objetivo específico al que se refieran.

Presentaciones del grupo y discusión plenaria (opcional)

- Pídales a todos los grupos que presenten al menos una tabla (ya llenada), así como las conclusiones derivadas de cada uno de sus objetivos que requieren tabulaciones cruzadas. Invite a otros grupos y a los facilitadores a hacer comentarios.

Módulo 25

MEDIDAS DE LA ASOCIACIÓN CON BASE EN EL RIESGO

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 25 MEDIDAS DE LA ASOCIACIÓN CON BASE EN EL RIESGO

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Definir* la incidencia, el riesgo, el riesgo relativo y la razón de probabilidades.
2. *Calcular* el riesgo relativo para poder hacer diseños de estudio apropiados (estudios comparativos transversales, estudios de cohorte, estudios de casos y controles y estudios experimentales).

I. Introducción.

II. Incidencia, riesgo y riesgo relativo.

III. Cálculo del riesgo relativo en diferentes diseños de estudio.

I. INTRODUCCIÓN

En la ISS, el objetivo de muchos estudios comparados (comparativos transversales, de casos y controles, de cohortes, experimentales y cuasi-experimentales) es comparar a las personas que contraen el problema o enfermedad en estudio entre quienes tienen un factor de riesgo (grupo expuesto), y quienes no lo tienen (grupo no expuesto). Cómo se haga esta comparación depende del diseño de estudio que se utilice.

Recuerde que en los *estudios de casos y controles* y en los estudios comparativos transversales (*módulo 9*), se selecciona un grupo de individuos que presente el problema en estudio (casos) y un grupo de control compuesto por individuos que no tienen el problema. Los dos grupos se comparan con respecto a la presencia o ausencia de factores de riesgo. En un *estudio de cohorte*, sólo se seleccionan sujetos sin el problema en estudio. Luego se dividen en dos grupos:

aquellos expuestos al factor de riesgo y los que no lo están; se les da seguimiento durante un periodo más largo y se mide en ambos grupos la aparición del problema. En un *estudio experimental*, se selecciona a individuos con un determinado problema. Un grupo se somete a una intervención por parte del investigador (esto es lo que los epidemiólogos llaman “factor de riesgo beneficioso”), mientras que el otro grupo no se somete a ninguna intervención. Después de un periodo de seguimiento adecuado, se mide la incidencia del problema y se comparan los dos grupos, esperando que la intervención haya solucionado por lo menos en parte el problema.

La comparación de un grupo expuesto al factor de riesgo y otro que no lo esté, o de casos y controles, permite al investigador determinar si el factor de riesgo potencial influye en el problema o no, y *en qué medida* contribuye efectivamente el problema.

Antes de poder medir qué tanto contribuye un determinado factor de riesgo a un problema, deben comprenderse los conceptos de incidencia, riesgo y riesgo relativo.

II. INCIDENCIA, PREVALENCIA, RIESGO Y RIESGO RELATIVO

Incidencia y tasa de incidencia

La *incidencia* es el número total de eventos o casos *nuevos* de un criterio definido (por ejemplo, una enfermedad) que se producen durante un determinado periodo en una población definida, que puede contraer la condición (enfermedad) de interés.

Nota: La población que puede contraer la condición (enfermedad) de interés también se conoce como población “de riesgo”.

Ejemplo 1:

El total de casos nuevos de tuberculosis en el distrito A, en el año 2000, fue de 273. La *incidencia* de la tuberculosis en el distrito A ese año, fue, por lo tanto, de 273.

La *tasa de incidencia* (incidencia acumulada) es el número total de nuevos eventos o casos de un criterio definido que se producen durante un periodo determinado dividido entre la población “de riesgo”.

Una tasa de incidencia se expresa generalmente por 1.000, por 10.000 o por 100.000 (u otro factor de 10) habitantes, para facilitar la comparación de las tasas en las diferentes comunidades.

Ejemplo 1 (continuación):

El distrito A tiene una población de 200,000 habitantes. La *tasa de incidencia* de la tuberculosis en el distrito A, en el año 2000, fue, por lo tanto, 273/200,000/año o 137/100,000/año. (Divida el numerador y el denominador entre 2 para expresar la tasa de incidencia por cada 100,000 habitantes al año, y redondee al número entero más próximo; es decir, 136.5 se convierte en 137).

La tasa de incidencia estima las probabilidades (o el riesgo) de que un individuo contraiga una enfermedad durante un determinado periodo.

Nota:

La *prevalencia* es el total de *los casos nuevos y los viejos*, independientemente de cuándo hayan ocurrido. La TASA DE PREVALENCIA (los casos nuevos y los viejos divididos entre la población total) es muy útil en la gestión, ya que da una indicación de los recursos necesarios para hacer frente al problema.

Ejemplo 1 (continuación):

Si el 31 de diciembre de 2000 se registran 360 pacientes entre nuevos y antiguos, la tasa de prevalencia es de 360/200.000 o 180/100.000 al año.

Al igual que las tasas de incidencia, las tasas de prevalencia pueden expresarse por cada 1.000, 10.000 o 100.000 habitantes.

Riesgo y riesgo relativo

El *riesgo* es lo mismo que la tasa de incidencia.

Ejemplo 1 (continuación):

El *riesgo* de contraer tuberculosis en el distrito A en el año 2000 fue de 137/100.000/año.

El riesgo puede no ser el mismo para los diversos subgrupos de la población. Mientras que el riesgo de contraer tuberculosis entre los agricultores podría ser de 100/100.000/al año, entre los mineros puede ser de 200/100.000/al año. En este ejemplo, los mineros son dos veces más propensos que los agricultores a contraer tuberculosis.

Puede concluirse, por lo tanto, que ser minero es un *factor de riesgo* para contraer tuberculosis y conlleva un *riesgo relativo* de 2.

Un *factor de riesgo* es cualquier factor cuya presencia se asocie con un aumento del riesgo de contraer una condición o una enfermedad.

Al determinar el riesgo relativo debemos considerar dos subgrupos de la población de estudio: el subgrupo en que el factor de riesgo está presente (expuestos) y aquel en que el factor de riesgo está ausente (no expuestos).

El *factor de riesgo relativo* es el riesgo de contraer la enfermedad en el grupo con el factor de riesgo (expuesto), dividido entre el riesgo de contraer la enfermedad en el grupo sin el factor de riesgo (no expuesto).

$$\text{Riesgo relativo (RR)} = \frac{\text{Tasa de incidencia (riesgo) entre aquellos con el factor de riesgo (expuestos)}}{\text{tasa incidencia (riesgo) entre aquellos sin el factor de riesgo (no expuestos)}}$$

Nota:

Si $RR = 1$, el riesgo de enfermedad es igual entre aquellos con el factor de riesgo y los que no lo tienen. Por lo tanto, no existe asociación entre el factor de riesgo y el problema o condición en estudio.

Si $RR > 1$, el riesgo de enfermedad es mayor entre aquellos con el factor de riesgo que entre los que no lo tienen. En este caso el factor se asocia con el problema o condición en estudio.

Si $RR < 1$, el riesgo de enfermedad es menor entre aquellos con el “factor de riesgo” que entre quienes no lo tienen. En este caso el factor de riesgo en realidad protege de –o reduce– el problema. (Por ejemplo, un “factor de riesgo” beneficioso tal como un Programa de educación sanitaria o alguna otra intervención, que ayuda a resolver –o a reducir– un determinado problema).

La identificación de un factor de riesgo *no* implica que entre éste y la condición haya una *relación causal*. Sin embargo, cuanto mayor es el riesgo relativo, más probabilidades hay de que el factor de riesgo sea causal y no debido a la casualidad o a un efecto de confusión.

III. CÁLCULO DEL RIESGO RELATIVO EN LOS DISTINTOS DISEÑOS DE ESTUDIO

1. Cálculo del riesgo relativo en estudios de cohorte y en estudios de intervención

En los estudios de cohorte y los estudios de intervención, la tasa de incidencia (riesgo) puede calcularse directamente. Esto se debe a que los resultados o las enfermedades o problemas en estudio se producen durante el estudio.

En estos estudios, se calculan las tasas de incidencia en los grupos expuestos y los no expuestos, y éstas se utilizan después para calcular el *riesgo relativo* mediante la fórmula presentada en la sección II de este módulo.

En un estudio de cohorte, los datos deben ingresarse en el formato expuesto en la *tabla 25.1*.

Tabla 25.1 Formato de la tabla para calcular el riesgo en los estudios generales de cohorte

Presencia del factor de riesgo	Presencia del problema		Total
	Sí	No	
Sí	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	a + b + c + d

En donde: a = Individuos con el factor de riesgo, que tienen también el problema en estudio
b = Individuos con el factor de riesgo, que no tienen el problema
c = Individuos sin el factor de riesgo, que tienen el problema
d = Individuos sin el factor de riesgo que no tienen el problema

De donde: a + b = todos los individuos que tienen el factor de riesgo
c + d = todos los individuos que no tienen el factor de riesgo

El riesgo de contraer la enfermedad entre aquellos con el factor de riesgo = $a/(a+b)$

El riesgo de contraer la enfermedad entre aquellos sin el factor de riesgo = $c/(c+d)$

El *riesgo relativo* va a ser entonces igual a $\frac{a/(a+b)}{c/(c+d)} = \frac{a/(c+d)}{c/(a+b)}$

Ejemplo 2:

En el país X se llevó a cabo un estudio para averiguar si el riesgo de tener diarrea en los niños menores de cinco años era diferente entre dos distritos cercanos (cada uno compuesto por un determinado número de poblados). Se sospechaba que el riesgo fuera mayor en el distrito B debido a que la comunidad utilizaba pozos desprotegidos, mientras que la población del distrito A utilizaba agua de perforaciones. El distrito A tenía una población de 10,000 habitantes, mientras que la población del distrito B era de 15,000 habitantes. La población de niños menores de cinco años era de 1,000 y 1,500 respectivamente. Se verificaron los registros hechos por los trabajadores de la salud de las poblaciones (TSP), responsables de los distritos, de las cuatro semanas anteriores. El distrito B había tenido 78 casos de diarrea mientras que el distrito A había tenido 50 casos.

¿El riesgo de tener diarrea es diferente entre estos dos distritos?

Como el distrito B tenía el factor de riesgo potencial (fuentes de agua desprotegidas), se esperaba que su población estuviera en un mayor riesgo de tener diarrea. Estos datos pueden presentarse ahora en una tabulación cruzada utilizando el formato de la tabla 25.1. El *riesgo* y el *riesgo relativo* pueden calcularse entonces con la fórmula comentada.

Tabla 25.2 Relación entre la exposición a agua insalubre y la diarrea en niños menores de 5 años

Exposición	Presencia de diarrea		Total
	Diarrea	No diarrea	
Agua insalubre (distrito B)	78 (5%)	1.422 (95%)	1.500 (100%)
Agua salubre (distrito A)	50 (5%)	950 (95%)	1.000 (100%)
Total	128 (5%)	2.372 (95%)	2.500 (100%)

Combine las *tablas 25.1 y 25.2*: $a = 78; b = 1,422; c = 50; d = 950$
 $a+b = 1,500; c+d = 1,000$

Riesgo de tener diarrea en el distrito B (agua insalubre) $= a/(a+b) = 78/1,500$

Riesgo de tener diarrea en el distrito A (sin factor de riesgo) $= c/(c+d) = 50/1,000$

El *riesgo relativo* va a ser, entonces, $\frac{78/1.500}{50/1,000} = \frac{78 \times 1.000}{50 \times 1.500} = 1,04$
igual a

El intervalo de confianza (IC) al 95% para esta RP –ya calculada para usted - el resultado va de 0.74 a 1.47.

Interpretación:

Los niños menores de 5 años del distrito B tienen < 1.04 veces más riesgo de tener diarrea que los del distrito A, pero este riesgo no es significativamente diferente puesto que el IC al 95% incluye el número 1. (Esta significación se va a evaluar más adelante, en los *módulos 28-30*, mediante la aplicación de pruebas estadísticas de significación.)

Nota:

El cálculo del riesgo relativo en los diseños de estudio experimentales, cuasi-experimentales y comparativos transversales basados en datos de incidencia es exactamente el mismo que en los estudios de cohorte. (Véanse en el anexo 25.1 ejemplos del cálculo de riesgos relativos en estos diseños).

2. Estimación del riesgo relativo en un estudio de casos y controles

En un estudio de casos y controles, por lo general no se puede calcular directamente el riesgo relativo, como en los estudios de incidencia comentados un poco antes.¹ En la mayoría de las situaciones, se emplea la razón de probabilidades (RP), que *estima* el riesgo relativo.

Para computar la RP, los datos deben presentarse en el formato expuesto en la *tabla 25.3*.

Tabla 25.3 Formato de la tabla para los estudios de casos y controles

Factor de riesgo	Casos	Controles	Total
Presente (+)	a	b	a + b
Ausente (-)	c	d	c + d
Total	a + c	b + d	a + b + c + d

Donde: a = Individuos con el problema (casos) que tienen el factor de riesgo
 b = Individuos sin el problema (controles) que tienen el factor de riesgo
 c = Individuos con el problema (casos) que no tienen el factor de riesgo
 d = Individuos sin el problema (controles) que no tienen el factor de riesgo

 RP = $\frac{ad}{bc}$

La RP se llama a veces *razón de productos cruzados*, pues se trata del producto de la casilla superior izquierda (a) y la casilla inferior derecha (d) o a por d, dividido entre el producto de la casilla superior derecha (b) y la casilla inferior izquierda (c), o b por d.

Una vez que calculamos la razón de probabilidades (RP), por lo general la llamamos riesgo relativo (RR) puesto que es esto lo que está estimando.

Ejemplo 3:

En Namibia se llevó a cabo un estudio de ISS de casos y controles para identificar los factores que contribuyen a la *mortalidad neonatal* precoz (primeros siete días de vida) en el hospital de maternidad de la capital, Windhoek (Muharukua *et*

1 En los “estudios de incidencia”, la incidencia de un determinado problema o resultado y el riesgo o el factor contribuyente pueden ambos ser observados; por lo tanto, el riesgo relativo puede calcularse directamente.

al. 1998). Por cada caso, se seleccionaron 5 controles. El tamaño de la muestra final fue de 290, de los cuales 49 eran recién nacidos que fallecieron entre el nacimiento y el séptimo día de vida (casos) y 241 neonatos que sobrevivieron a los primeros siete días de vida (controles). Entre los posibles factores de riesgo evaluados estuvo el bajo peso al nacer (menos de 2.500 g).

De los 281 recién nacidos de quienes se dispuso de información sobre su bajo peso al nacer (BPN), 44 eran casos y 237 controles. De los 44 casos, 28 tuvieron bajo peso al nacer, mientras que de los 237 controles, 29 tuvieron bajo peso.

Estos datos, presentados en el formato de tabulación cruzada expuesto en la *tabla 25.3*, se muestran en la *tabla 25.4*.

Tabla 25.4 Relación entre los niños prematuros y los neonatos fallecidos, en un estudio de casos y controles en Namibia

Bajo peso al nacer	Muertes neonatales	Controles	Total
Sí	28 (64%)	29 (12%)	57
No	16 (36%)	208 (88%)	224
Total	44 (100%)	237 (100%)	281

Donde $a = 28$; $b = 29$; $c = 16$; $d = 208$

La fórmula de la razón de probabilidades (RP) es:

$$\text{RP} = \text{RR} = \frac{\frac{ad}{bc}}{\frac{28 \times 208}{29 \times 11}} = 12,55$$

El IC al 95% es de 5,69 a 28,00. (Éste se calculó en EpiTable, usando el Curso a distancia del programa Epi.info versión 6.04c, como se muestra en el “Computer Companion”; Vol. II, parte 3 de la Serie de capacitación para la ISS, la OMS / AFRO, Harare, 1996). (Véanse en el módulo 27 las fórmulas para calcular el IC al 95%).

Interpretación:

En el hospital de maternidad de Windhoek, el riesgo de que un neonato con bajo peso al nacer muera en los primeros siete días de la vida es 12,55 veces mayor que la de los recién nacidos con peso normal. El bajo peso al nacer es, por lo tanto, un factor de riesgo muy poderoso para la muerte neonatal.

Note usted que los porcentajes de las tabulaciones cruzadas ya le dieron una pista de que en la *tabla 25.2* la RP/RR sería baja, mientras que en la *tabla 25.4* usted pudo suponer que ésta sería muy alta. Estos dos ejemplos se proporcionan con fines didácticos. En realidad, ni siquiera tuvo usted que molestarse en calcular la razón de probabilidades para la *tabla 25.2*.

3. Cálculo del riesgo relativo en un estudio comparativo transversal (investigación de la prevalencia)

Al igual que en un estudio de casos y controles, en un estudio comparativo transversal de *prevalencia* la incidencia no suele calcularse directamente. La medida de asociación en estos estudios se llama razón de probabilidades de prevalencia (RPP). La RPP se calcula de la misma manera que la razón de prevalencia en un estudio de casos y controles.

Ejemplo 4:

En Botsuana se llevó a cabo un estudio comparativo transversal para determinar la magnitud del problema de los *embarazos en la adolescencia* e identificar los factores contribuyentes. Los investigadores tomaron al azar una muestra de 400 adolescentes, y encontraron que 23% de ellas habían estado embarazadas. Querían evaluar, entre otras cosas, si las adolescentes que habían recibido apoyo de alguna organización (es decir, educación entre iguales) tenían una menor probabilidad de quedar embarazadas que las que no habían recibido apoyo.

De los datos recogidos, sólo 14 de las 90 adolescentes que habían tenido un embarazo habían recibido apoyo de alguna organización, mientras que 86 de las 310 adolescentes que nunca se habían embarazado habían recibido apoyo. Estos datos se exponen en la *tabla 25.5*.

Tabla 25.5 Relación entre el apoyo de las organizaciones y el embarazo en la adolescencia en Botsuana

Apoyo de las organizaciones	Embarazo		Total
	Sí	No	
No	76 (25%)	224 (75%)	300 (100%)
Sí	14 (14%)	86 (86%)	100 (100%)
Total	90 (23%)	310 (77%)	400 (100%)

En donde a = 76; b = 224; c = 14; d = 86

(La prevalencia) razón de probabilidades (RP) = $\frac{a \times d}{b \times c}$

RP = RR = $\frac{28 \times 208}{29 \times 11} = 2,08$

El IC al 95% es de 1,07 a 4,11 (calculado en EpiTable, usando el Curso a distancia de Epi.info versión 6.04c, como se muestra en el “Computer Companion”, Vol. II, parte 3 de la Serie de capacitación para la ISS).

Interpretación:

Las adolescentes que no recibieron apoyo de ninguna organización tuvieron 2 veces más probabilidades de embarazarse que aquellas que recibieron apoyo. Note que este hallazgo tiene significación tanto práctica como estadística.

Advertencia

Cuando se encuentra una asociación entre un factor de riesgo potencial y el problema, puede no tratarse de una relación real. Un hallazgo semejante puede tener diversas razones, una de las cuales es que puede haber otros factores de riesgo (factores de confusión) que dan lugar a esa explicación. Estos otros factores de riesgo deben, por lo tanto, tomarse en cuenta. El *módulo 26* trata el tema del fenómeno de confusión, incluyendo la manera de evaluarla y controlarla.

TRABAJO EN EQUIPO

Si usted hizo un estudio comparado (comparativo transversal, de cohorte, de casos y controles o (cuasi) –experimental):

- Analice sus tablas cruzadas de dos por dos y seleccione aquéllas en que los porcentajes indiquen que podría haber una asociación entre dos variables.
- Calcule la razón de probabilidades (RP) para cada tabla.
- Interprete los resultados y escriba la interpretación debajo de cada tabla.

El intervalo de confianza (IC) al 95% para esta RP es de 1.64 a 3.98. (Éste se calculó para usted en EpiTable, usando el Curso a distancia del programa Epi.info versión 6.04c, como se muestra en el “Computer Companion”, Vol. II, parte 3 de la Serie de capacitación para la ISS).

Interpretación:

El paquete de capacitación para la salud funciona. El riesgo de diarrea en menores de cinco años después de la intervención es ahora mucho más alto (2.29 veces) en el distrito B que en el A. El hallazgo es importante en la práctica (el riesgo de diarrea es dos veces mayor en el grupo en donde no hubo intervención) y es estadísticamente significativo (95% de los niveles de confianza no incluyen el 1). La prueba estadística formal se trata en los *módulos 29 y 30*. Recuerde que el riesgo de diarrea no fue significativamente diferente antes de la intervención (*tabla 25.2*).

Cálculo del riesgo relativo en un estudio comparativo transversal (estudio de la incidencia)

La tasa de incidencia también se puede calcular directamente en un estudio comparativo transversal donde se identifiquen los casos incidentes (problema). Esto puede utilizarse después para calcular el riesgo relativo, como ya se demostró (*tabla 25.2*).

Los datos deben verse, sin embargo, en un formato de tabla como el expuesto en la *tabla 25.7*.

Tabla 25.7 Formato de la tabla para un estudio transversal (resultados incidentes)

Factor de riesgo	Incidente Sí	Incidente No	Total
Presente (+)	a	b	a + b
Ausente (-)	c	d	c + d
Total	a + c	b + d	a + b + c + d

Donde:

- a = Individuos con el factor de riesgo que también tienen el problema o el incidente en estudio
- b = Individuos con el factor de riesgo que no tienen el problema
- c = Individuos sin el factor de riesgo que también tienen el problema
- d = Individuos sin el factor de riesgo que no tienen el problema

Puede, entonces, aplicarse la fórmula para calcular el riesgo relativo.

Incidenia en los expuestos = $a/(a+b)$

Incidencia en los no-expuestos = $c/(c+d)$

El riesgo relativo puede escribirse también como: $RR = \frac{a/(c + d)}{c/(a + b)} = \frac{a \times (c + d)}{c \times (a + b)}$

Ejemplo 5

En un estudio sobre los factores que contribuyen al retraso en la búsqueda de atención perinatal, llevado a cabo en el distrito de Machinga, Malawi, en 1997 (Tamaona *et al.*, 1997), 88% de las 97 madres entrevistadas postergaron la búsqueda de atención perinatal (hasta después del tercer mes de embarazo). Para averiguar si la distancia de las instalaciones de salud era un factor de riesgo para este atraso en la búsqueda de atención perinatal, se separó a las madres en dos grupos: las que vivían a 10 km o menos del centro de salud (cerca) y las que vivían a más de 10 km de distancia (lejos). Se encontró que 60 de las 62 madres que vivían lejos del centro de salud retrasaron la búsqueda de atención perinatal, en comparación con 28 de las 35 madres que vivían cerca.

Los datos se exponen en la *tabla 25.8* con el mismo formato indicado en la *tabla 25.7*. La incidencia (el problema) es el retraso en la búsqueda de atención perinatal, mientras que el factor de riesgo es que viven lejos (> 10 km).

Tabla 25.8 Relación entre la distancia y el retraso en la búsqueda de atención perinatal en el distrito de Machinga, Malawi

Distancia de las instalaciones de salud	Búsqueda de atención perinatal		Total
	Hubo retraso	No hubo retraso	
> 10 km	60	2	62
0 – 10 km	28	7	35
Total	88	9	97

De las madres que vivían lejos, 96,77% se retrasaron. Esto es lo mismo que

$$60/62 = 96,77/100 = 0,9677$$

Por otra parte, 80,0% de las que vivían cerca se retrasaron, es decir

$$28/35 = 80,00/100 = 0,80$$

$$RR = \frac{0,9677}{0,80} = 1,21$$

El del IC al 95% para esta RP es de 1,02 a 1,44. (Éste se calculó usando Epi-Info en el programa Epi-Info versión 6.04c, como se muestra en el “Computer Companion,” Vol. II, parte 3 de la Serie de capacitación para la ISS).

Interpretación:

Las mujeres que viven a más de 10 km de un centro de salud en el distrito de Machinga, Malawi son, por lo tanto, 1.21 veces más propensas a retrasar la presentación de la atención perinatal que las que viven a 10 km o menos de las instalaciones de salud más cercanas. La diferencia es significativa estadísticamente, aunque no muy importante (sólo 1.2 veces más probabilidades de retraso). La mejora del acceso a los servicios podría no cambiar mucho el retraso en la búsqueda de atención perinatal. Debe de haber otros factores, además de la distancia, que hacen que las madres se mantengan alejadas de los servicios de salud.

Módulo 25 LAS MEDIDAS DEL RIESGO Y LA ASOCIACIÓN

Tiempo estimado y recursos didácticos

1 ½ hora	Introducción y discusión
2 horas	Trabajo en equipo

Introducción y discusión

Es importante que los participantes entiendan el concepto de riesgo. Procure cerciorarse de que también tengan claro cómo aplicar este concepto en cualquier diseño de estudio. Permita que en la sesión plenaria calculen algunas razones de posibilidades con sus propios datos.

Subraye la importancia del diseño de estudio cuasi-experimental, y cómo se puede utilizar para evaluar los programas de una manera científica.

El equipo de facilitadores debe proporcionar apoyo informático para mejorar la profundidad de análisis, ya que el análisis manual de los datos puede dar lugar a una visión inadecuada de la contribución de los factores de riesgo al problema en estudio.

Módulo 26

MANEJO DE LAS VARIABLES DE CONFUCIÓN

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 26 MANEJO DE LAS VARIABLES DE CONFUSIÓN

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Explicar* diferentes maneras de manejar el fenómeno de confusión en la etapa de diseño y análisis de un estudio.
2. *Evaluar* si una asociación entre dos variables puede estar influida por otra variable de confusión/factor de riesgo.
3. *Calcular* la asociación de una manera que considere el efecto potencial del fenómeno de confusión por parte de otra variable/factor de riesgo.

I. Introducción.

II. Naturaleza del fenómeno de confusión.

III. Métodos para controlar el fenómeno de confusión en el análisis.

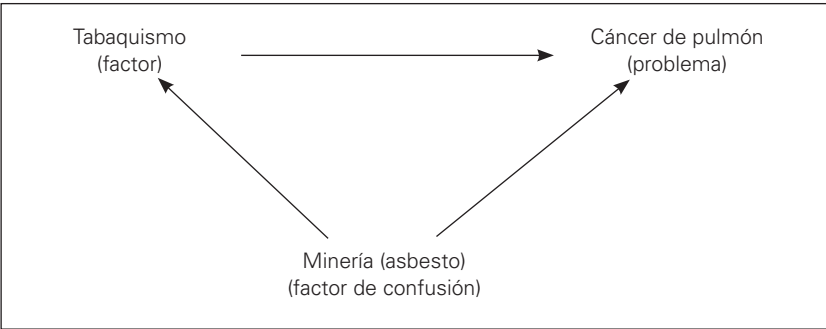
I. INTRODUCCIÓN

En el *módulo 25* calculamos el riesgo y luego medimos la asociación entre los factores de riesgo y el problema objeto de estudio (resultado). Sin embargo, puede ser que la asociación observada se deba parcial o totalmente a otro factor de riesgo, y no al que se está considerando. Si esta “distorsión” –debida a un segundo factor de riesgo–, no se toma en cuenta, no podremos saber si la primera asociación realmente existe ni si es tan fuerte como se calculó.

Por ejemplo (1), se sabe que el tabaquismo es un factor de riesgo para el cáncer de pulmón. Sin embargo, en la región Z, muchos hombres trabajaban en las minas de asbesto locales y estuvieron, por consiguiente, expuestos al asbesto, que es un sabido riesgo para el cáncer de pulmón. Se sabe también que por la tensión a la que están expuestos los mineros, tienden a fumar más, especial-

mente cuando trabajan bajo tierra. Puesto que el tabaquismo se relaciona con el cáncer de pulmón, la minería se relaciona con el tabaquismo y con el cáncer de pulmón. Hay, por lo tanto, una relación triangular entre el tabaquismo, la minería y el cáncer de pulmón, como se muestra en la *ilustración 26.1*.

Ilustración 26.1 Interrelación entre el tabaquismo (factor), la minería (factor de confusión) y el cáncer de pulmón (problema) en un estudio de cohorte



En este módulo vemos cómo puede descartarse la influencia de las variables de confusión y cómo calcular la medida en que el factor de riesgo original —en el que un investigador quiere concentrarse (fumar en el ejemplo 1), contribuye al problema objeto de la investigación (cáncer de pulmón).

II. NATURALEZA DEL FENÓMENO DE CONFUSIÓN

¿Qué es un factor de confusión?

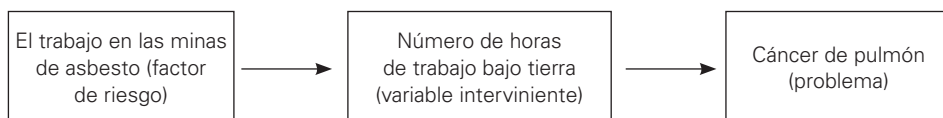
Un factor de confusión es una variable independiente que distorsiona la asociación entre otra variable independiente y el problema objeto de estudio, ya que está relacionado con ambos.

Para que una variable sea un factor de confusión, debe estar asociada con el primer factor de riesgo (el tabaquismo, en el ejemplo 1) y debe ser un factor de riesgo independiente del problema (cáncer de pulmón). Por otra parte: una variable que tan solo es un vínculo entre el factor de riesgo que estamos buscando y el problema, no es un factor de confusión potencial.

Por ejemplo, si pudiéramos relacionar el número de horas que los mineros están bajo tierra con el cáncer de pulmón (véase la ilustración 26.2), éste simplemente precisaría los factores de riesgo del trabajo en las minas. Por consiguiente, no

empañosaría la asociación entre el trabajo en las minas de asbesto y el cáncer de pulmón. (Véase también el ejemplo del módulo 23, Sección III.)

Ilustración 26.2 Interrelación entre un factor, una variable interviniente y el problema o resultado



Para que un factor pueda verse como una variable de confusión potencial, debe haber una relación triangular entre el primer factor de riesgo, el factor de confusión potencial y el problema objeto de investigación, como se muestra en la *ilustración 26.1*.

¿Cuál es el efecto del fenómeno de confusión?

El fenómeno (o efecto) de confusión puede resultar de que la asociación entre un factor de riesgo y el resultado parezca menor (se subestime) o parezca mayor (se sobre-estime) de lo que es. Puede incluso cambiar la dirección del efecto observado y dar lugar a que un factor perjudicial parezca ser beneficioso y viceversa.

Reconsideremos el efecto de trabajar en las minas de asbesto en la relación entre el tabaquismo y el cáncer de pulmón.

Ejemplo 1 (continuación)

Se encontró que en la región Z, el riesgo de que un fumador contrajera cáncer de pulmón era 10 veces el de un no fumador. Cuando los investigadores descartaron el efecto de confusión de trabajar en las minas de asbesto, descubrieron que los fumadores tenían sólo 7 veces más probabilidades de contraer cáncer de pulmón. La verdadera relación entre el tabaquismo y el cáncer de pulmón fue, por lo tanto, 30% menor que cuando no se había descartado el factor de confusión.

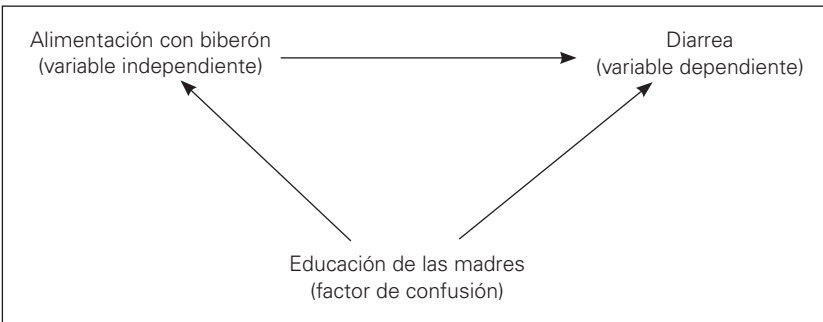
En el *módulo 8*, donde introdujimos el concepto de variable de confusión, dimos un ejemplo de cómo el efecto de la práctica de que una madre alimente a su niño con biberón sobre la diarrea del niño, puede confundirse por el nivel educativo de la madre. Ahora vamos a retirar el efecto de la educación de la madre, para poder descubrir la verdadera relación entre la alimentación con biberón y la diarrea.

Ejemplo 2

En un estudio comparativo transversal utilizando casos prevalentes de niños que nunca habían tenido diarrea (comentado en el *módulo 24*, sección V), se supuso que los lactantes alimentados con biberón son más propensos a tener diarrea que los lactantes amamantados. Se sabe que las madres sin educación tienden menos a darles biberón a sus niños, pero aun así sus hijos son más propensos a tener diarrea que los de sus contrapartes más educadas. La educación de las madres es, por consiguiente, un posible factor de confusión en la relación entre la alimentación con biberón y la diarrea (véase la ilustración 26.3).

En este estudio, se descubrió que los lactantes alimentados con biberón tenían menos probabilidades (0.8 veces o una $RP = 0.8$) de tener diarrea que aquéllos alimentados con leche materna. Esto sugería que ¡la alimentación con biberón *protegía* a los lactantes de la diarrea! Sin embargo, cuando se tomó en cuenta la educación de la madre, se descubrió que los lactantes alimentados con biberón tenían en realidad 1.2 veces más probabilidades de tener diarrea que los lactantes amamantados. Esta asociación se había distorsionado por el efecto de la educación materna. La verdadera relación entre la alimentación con biberón y la diarrea era en realidad 50% mayor que la distorsionada (por el efecto de confusión).

Ilustración 26.3 Interrelación entre la alimentación con biberón (riesgo), la educación de la madre (factor de confusión) y la diarrea (problema) en un estudio de casos y controles



Cómo identificar los factores de confusión

Puesto que los factores de confusión pueden distorsionar la verdadera relación entre los factores de riesgo y el problema en estudio, es fundamental considerarlos en el *diseño del análisis de datos*. Para controlar las variables de confusión en la etapa del análisis de los datos, deben recogerse durante la propia recolección los datos sobre los factores de confusión; de otro modo no podrá descartarse la distorsión debida a dichos factores.

El diagrama de análisis del problema (*módulo 4*) ofrece mayor información sobre algunas posibles variables de confusión, ya que permite visualizar las relaciones triangulares que acabamos de comentar. Deben considerarse las variables de confusión que se encuentren en la bibliografía, además de las que puedan identificarse durante la lluvia de ideas sobre los posibles factores de riesgo y sus interrelaciones. Las *variables contextuales* tales como la edad, la educación y el estatus económico son notorios factores de confusión, que están relacionados con diversos aspectos de la vida (véase el módulo 8). A menudo están incluso interrelacionados; por ejemplo, tanto la educación como la edad pueden contribuir a la situación económica.

Cuando se observan hallazgos inusuales o inesperados durante el análisis de los datos, pueden identificarse otros posibles factores de confusión. Así, cuando las relaciones son más fuertes, más débiles o incluso contrarias a lo que esperaba (por ejemplo, beneficiosas en vez de perjudiciales) a partir de la bibliografía o de la propia experiencia, esto debe tomarse como aviso para buscar factores de confusión.

EJERCICIO

Utilice su diagrama de análisis del problema y sus hallazgos para identificar los posibles factores de confusión.

III. MÉTODOS PARA CONTROLAR EL FENÓMENO DE CONFUSION EN EL ANÁLISIS DE LOS DATOS

El control del fenómeno de confusión en el análisis de datos se logra mediante un análisis estratificado o un análisis multivariado. En el primero, el objetivo es comparar los grupos de caso y de control en categorías homogéneas respecto de la variable de confusión. El análisis multivariado, en cambio, implica el uso de un modelo matemático y permite que el investigador controle simultáneamente todas las variables de confusión, mientras busca la contribución de cada factor de riesgo para el problema. No describimos aquí el análisis multivariado, pero quienes estén interesados pueden consultar manuales epidemiológicos o estadísticos. (Véanse las referencias en los módulos 9 y 28). Cuando, al diseñar su estudio, los investigadores prevén algunas variables de confusión —en la relación del problema en estudio y un factor contribuyente— pueden aparear los casos y los controles de estos “posibles factores de confusión”. Los datos apareados requieren un análisis específico (véase la parte 2 de esta sección).

1. Análisis estratificado

Los pasos para controlar el fenómeno de confusión mediante el análisis estratificado son:

- a. Calcular el riesgo relativo (RR) o la razón de probabilidades (RP, que es una estimación del RR, véase el módulo 25), sin estratificar (RR crudo o RP cruda)
- b. Estratificar por la variable de confusión
- c. Calcular el RR ajustado (o RP)
- d. Comparar el RR crudo o RP con el RR ajustado o la RP

Si la estimación ajustada (RRa o RPa) es igual a la que no está ajustada (RR o RP), entonces no hay efecto de confusión. Si son *diferentes*, entonces sí lo hay. Pero a la pregunta, de qué tan grande debe ser la diferencia, los epidemiólogos han acordado que:

Si el *riesgo relativo crudo* difiere del *riesgo relativo ajustado* en 10% o más, sí hay un efecto de confusión importante. El RR ajustado debe entonces calcularse, estratificando la variable de confusión.

Puede hacerse el cálculo del IC al 95% (y la prueba de significación formal) para medir la significación de la asociación entre el factor de riesgo y el problema en los diferentes estratos.

(1) Análisis estratificado para controlar el fenómeno de confusión en un estudio de incidencia¹

El procedimiento para controlar el fenómeno de confusión es el mismo en todos los estudios de incidencia, incluyendo los estudios de casos y controles y los de cohortes, los comparativos transversales y los experimentales.

Ejemplo 3:

En el estudio de intervención presentado en el *módulo 25* (ejemplo 2), se sospechaba que la edad podía ser un factor de confusión. Se descubrió que los niños del distrito A eran por lo general mayores que los del distrito B (menores de 2 años versus 2 años y mayores) y los niños mayores eran más propensos a la diarrea. Era necesario evaluar, por consiguiente, si la edad tenía un efecto de confusión en la relación entre la intervención de educación para la salud y el episodio de diarrea, y dividir la *tabla 25.2* para los niños <2 años y los niños ≥ 2 años.

De los 1,500 niños del distrito B, 840 eran <2 años y 660 tenían 2 años o más. De los 1,000 niños del distrito A, 210 eran <2 años y 790 tenían 2 años o más.

De los 840 niños menores de dos años del distrito B, 40 tenían diarrea, en comparación con 10 de los 210 niños menores de dos años del distrito A. Entre los

¹ En los “estudios de incidencia”, puede observarse tanto la incidencia de un determinado problema o resultado como el riesgo o el factor contribuyente; por lo tanto el riesgo relativo puede calcularse directamente.

niños de 2 años y más, 15 de 660 tuvieron diarrea en el distrito B, frente al 6 de 790 en el distrito A (véanse las tablas 26.1 y 26.2).

Estas tablas muestran ahora los datos *estratificados por edad* para evaluar el efecto de confusión. Ya se había encontrado en el ejemplo 2 (anexo 25.1 del *módulo 25*), que el RR crudo era de 2.29, es decir, el riesgo de diarrea en el distrito donde no hubo intervención fue de poco más del doble que en el distrito donde sí la hubo, lo que sugería que el paquete de educación para la salud efectivamente funcionaba. El RR ajustado es de 1.51. (Véanse las fórmulas en el anexo 26.1. Para hacer los cálculos también puede emplearse Epi-Info en el programa Epi-Info 6.04, como se indica en el “Computer Companion,” Vol. II, parte 3 de la Serie de capacitación para la ISS).

Tabla 26.1 Riesgo de tener diarrea en los niños menores de 2 años por área de intervención (post intervención)

Exposición	Diarrea	No diarrea	Total
No hubo intervención (distrito B)	40	800	840
Hubo intervención (distrito A)	10	200	210
Total	50	1.000	1.050

Tabla 26.2 Riesgo de tener diarrea en los niños de 2 años y mayores por área de intervención (post intervención)

Exposición	Diarrea	No diarrea	Total
No hubo intervención (distrito B)	15	645	660
Hubo intervención (distrito A)	6	784	790
Total	21	1.429	1.450

¿El RR crudo (RR) difiere del RR ajustado (RRa)?

2,29 es diferente de 1,51.

¿La diferencia es mayor de 10%?

$$RR - RRa = 2,29 - 1,51 = 0,78$$

La diferencia entre RR y RRa como un porcentaje del RR es, por lo tanto:

$$\frac{0,78 \times 100}{2,29} = 34\%$$

Dado que la diferencia entre el RRa y el RR está por encima de 10% (en realidad es de 34%), hay un considerable efecto de confusión. Por consiguiente debe usarse el RRa; el RR está distorsionado por el efecto de la edad.

El 95% del IC calculado en EpiTable (programa Epi Info 6,04c) va de 0.89 a 2,55.

Interpretación:

La edad confundió (distorsionó) la relación entre la intervención y la aparición de diarrea, dando lugar a una sobre-estimación del RR de 2.29, cuando la relación es realmente de sólo 1.51. *La intervención realizada en el distrito A no parece haber funcionado*, pues aunque el riesgo de diarrea entre los niños menores de cinco años en el distrito donde hubo intervención es 1.51 veces el riesgo en el distrito donde no la hubo (distrito B), esto también puede deberse a la casualidad (95 % del IC incluye la cifra 1).

(2) Análisis estratificado para controlar el fenómeno de confusión en un estudio de casos y controles

El principio del control del fenómeno de confusión empleado para la estratificación es el mismo que para los estudios de casos y controles y los estudios comparativos transversales que utilizan los resultados de la prevalencia, puesto que ambos emplean la razón de probabilidades (RP de la prevalencia en este último) para estimar el riesgo relativo.

Ejemplo 4:

En el estudio de casos y controles realizado en el hospital de Windhoek, en Namibia, para encontrar los factores que contribuyen a la mortalidad neonatal precoz, ya mencionado en el ejemplo 3 del *módulo 25* (Muharukua *et al.*, 1998), se sospechó que la prematuridad tuvo un efecto de confusión en la relación entre el bajo peso al nacer (BPN) y la muerte neonatal.

Si se piensa bien, un neonato prematuro tiene más probabilidades de morir que un niño que nace a término, y un neonato prematuro tiene también una mayor probabilidad de pesar menos de 2.500 g (BPN). Es por estos motivos que se sospechaba que la prematuridad podría haber tenido un efecto de confusión en la relación entre el bajo peso al nacer y la muerte neonatal.

Ya se había descubierto que los neonatos con bajo peso al nacer tenían 12.55 veces más probabilidades de morir que los neonatos nacidos con peso normal. Este hallazgo fue significativo estadísticamente (95% del IC va de 5.69 a 28.00). Es posible, no obstante, que ésta no fuera la relación real, puesto que la prematuridad puede haber tenido en ella un efecto de confusión (distorsión). Es, pues, necesario estratificar los datos por el posible factor de confusión: la prematuridad. A continuación se exponen los datos que pueden utilizarse para obtener los dos estratos.

De los 281 neonatos (de los que se obtuvo datos sobre la prematuridad y el peso al nacer), 50 fueron prematuros mientras que 231 nacieron a término. Entre los prematuros, 27 de los 28 que murieron tenían bajo peso al nacer en comparación con 17 de los 22 controles. Para los neonatos nacidos a término, 1 de los 16 que murieron tenía bajo peso al nacer en comparación con 12 de los 215 controles. Las *tablas 26.3 y 26.4 dan la estratificación por prematuridad*, empleando estos datos.

Tabla 26.3 Relación entre el bajo peso al nacer y la muerte neonatal entre neonatos prematuros en el hospital de Windhoek

Peso bajo al nacer	Muertes neonatales	Controles	Total
Sí	27	17	44
No	1	5	6
Total	28	22	50

Tabla 26.4 Relación entre el bajo peso al nacer y la muerte neonatal entre los neonatos nacidos a término en el hospital Windhoek

Bajo peso al nacer	Muertes neonatales	Controles	Total
Sí	1	12	13
No	15	203	218
Total	16	215	231

La RP ajustada (RPa) es 3.20 (calculada mediante la fórmula expuesta en el *anexo 26.1* o empleando Epi table en el programa Epi Info 6.04c).

Ahora debemos preguntarnos si la RP es distinta de la RPa

Sí, lo es (¡12,55 versus 3,20!).

¿Por cuánto?

La relación ajustada de RPa = 3,20 es 75% menor que la del efecto de confusión (RP = 12,55).

Interpretación:

La prematuridad es un factor de confusión muy fuerte de la relación entre el bajo peso al nacer y la muerte neonatal. Si el efecto de confusión debido a la prematuridad no se hubiera tomado en cuenta, los investigadores podrían haber concluido que el bajo peso al nacer aumenta 12 veces el riesgo de muerte neonatal temprana, un hallazgo práctico y estadístico aparentemente muy importante. Pero que no es, sin embargo, real, puesto que la mayor parte de este alto riesgo no se debió al bajo peso al nacer en sí. Al descartar la contribución de la prematuridad (controlando los efectos de confusión de la prematuridad) se hace

evidente que el bajo peso al nacer aumenta el riesgo de muerte de un neonato sólo 3 veces ($RP_a = 3,20$, el 95% del IC va de 0,88 a 11,56), hallazgo que quizás no sea estadísticamente significativo (95% del IC incluye la cifra 1).

Puede concluirse correctamente, entonces, que *el bajo peso al nacer no es de suyo un factor de riesgo en la muerte neonatal*, sino que tiene un efecto sobre todo cuando hay prematuridad. *Es, por consiguiente, ésta la que, de tratarse, puede reducir la incidencia de mortalidad neonatal.*

Explicaciones alternativas de los hallazgos que no son significativos

Si se descarta el efecto de confusión como posible explicación de los resultados y se encuentra una asociación que no es estadísticamente significativa, (mediante pruebas de intervalos de confianza o pruebas formales de significación; véanse los módulos 29 y 30), hay dos explicaciones alternativas.

- i. El estudio no tuvo el poder suficiente para mostrar la diferencia, incluso si ésta existe en la población. Esto significa que los individuos eran muy pocos, lo que puede verse a partir de los *intervalos de confianza*; si los IC son amplios (por ejemplo, de 0,8 a 50), ello sugiere un problema de poder. Podría haber la necesidad de hacer otro estudio para descartar la posibilidad del fenómeno confusión. Este problema debe prevenirse durante la elaboración de la propuesta de investigación, calculando un tamaño de la muestra que garantice la viabilidad del estudio (véase el módulo 11).
- ii. Este es un verdadero hallazgo y no hay realmente ninguna asociación o relación entre el factor de sospecha y el problema. *Los intervalos de confianza estrechos* (por ejemplo, de 0,85 a 4,55) indican que éste podría ser el caso.

2. Análisis apareado

Para tratar el fenómeno de confusión en algunos estudios de casos y controles, se aparea un control a cada caso. Este apareamiento debe incluirse en el diseño del estudio; diseño que hay que tener en cuenta al hacer el análisis. Sin embargo, si ya se ha realizado un estudio de frecuencia de controles y casos apareados, entonces debe hacerse, en su lugar, un análisis estratificado, como se hizo en las tablas 26.3 y 26.4.

Cálculo del riesgo relativo (RP) en los estudios de casos y controles apareados

En un análisis de casos y controles apareados es importante considerar el número de pares y no el número real de personas del estudio.

Ejemplo:

Se llevó a cabo un estudio de casos y controles para determinar las causas de un brote de cólera en uno de los barrios bajos de Bombay, en India. Por cada caso de cólera (confirmado bacteriológicamente) se buscó un control del mismo sexo, nacido en la misma década, y que viviera en el mismo barrio. La *tabla 26.5* presenta los resultados:

Tabla 26.5 Fuente de agua potable por par de paciente y control durante los 5 días previos a la enfermedad

Controles sanos	Casos de cólera		Total de pares
	Pozo superficial	Agua de grifo	
Pozo superficial	12	3	15
Agua de grifo	30	31	61
Total	42	34	76

Se descubrió que en 30 pares, los casos tomaban el agua de pozos superficiales, cosa que no hacían los controles, mientras que en sólo 3 pares, los controles tomaban el agua de pozos superficiales, mientras que los casos no lo hacían.

Por lo tanto, el riesgo relativo se calcula como:

Riesgo relativo = $\frac{30}{3} = 10$

En otras palabras, quienes sacaban agua de los pozos superficiales tenían 10 veces más probabilidades de contraer el cólera que los que obtenían el agua de grifos.

Los pasos para hacer este tipo de análisis general se describen a continuación:

Paso 1. Prepare una tabla:

Distribuya los pares en las 4 celdas de la tabla, según si

- a) tanto el caso como el control tienen el factor de riesgo (q)
- b) el control tiene el factor de riesgo y el caso no lo tiene (r)
- c) el control no tiene el factor de riesgo mientras que el caso sí lo tiene (s)
- d) Ni el control ni el caso tienen el factor de riesgo (t)

Controles	Casos	
	Factor de riesgo (+)	Factor de riesgo (-)
Factor de riesgo (+)	q	r
Factor de riesgo (-)	s	t

Paso 2. Revise si hay alguna asociación entre el factor de riesgo y la condición o enfermedad. Esto se hace comparando r y s en la tabla. Haga una prueba X^2 de McNemar para determinar si la asociación es estadísticamente significativa, es decir, si no se debe a la variación de la muestra (vea el módulo 30).

Paso 3. Calcule el riesgo relativo mediante la siguiente fórmula:

$$\text{Riesgo relativo} = \frac{s}{r} = \frac{30}{3} = 10$$

Trabajo en equipo

Si usted hizo un estudio comparativo (comparativo transversal, de cohorte, de casos y controles o (cuasi) experimental):

Identifique y enumere todas las variables de confusión potenciales y las relaciones sobre las que éstas pudieran tener un efecto de confusión.

Use la estratificación para controlar el efecto de confusión sobre todas las relaciones apropiadas en que sea necesario (si no está familiarizado con el programa Epi Info, su facilitador lo puede ayudar a conseguir las listas impresas de Epi table).

Re-interprete ahora cada una de las relaciones, y escriba un enunciado para cada una de ellas.

Anexo 26.1

Fórmula para calcular el riesgo relativo ajustado

En toda la fórmula, $T = a+b+c+d$ = total de los sujetos de estudio para el estrato o nivel.

Estudio de cohorte:

$$aRR = \frac{\Sigma(a \times \{c + d\}) / T}{\Sigma(c \times \{a + b\}) / T}$$

Para los datos estratificados del ejemplo (que se encuentran en las *tablas 26.1 y 26.2*):

$$\begin{aligned} RRa &= \frac{[(40 \times \{10+200\})/1.050] + [(15 \times \{6 + 784\})/1.450]}{(10 \times \{40 + 800\})/1.050 + [(6 \times \{15 + 645\})/1.450]} \\ &= \frac{(40 \times 210)/1,050 + [(15 \times 790)/1,450]}{[(10 \times 840)/1,050] + [(6 \times 660)/1,450]} \\ &= \frac{8 + 8.172}{8 + 2.731} = \frac{16.172}{10.731} = 1,51 \end{aligned}$$

Note que ésta es la misma respuesta que se obtuvo utilizando EpiTable en el programa Epi Info versión 6.04.

Estudio de casos y controles (tablas 26.3 y 26.4):

$$RPa = \frac{\Sigma(ad/T)}{\Sigma(bc/T)}$$

Por ejemplo,

$$\begin{aligned} RPa &= \frac{(27 \times 5)/50 + (1 \times 203)/231}{(17 \times 1)/50 + (12 \times 15)/231} = \frac{135/50 + 203/231}{17/50 + 180/231} \\ &= \frac{2,70 + 0,879}{8 + 0,779} = \frac{3.579}{1.119} = 3,20 \end{aligned}$$

Note, de nuevo, que esta respuesta concuerda enteramente con la que se calculó empleando EpiTable en el programa Epi Info versión 6.04.

**Módulo 26 MANEJO DEL FENÓMENO DE CONFUSIÓN
EN LA INVESTIGACIÓN**

Tiempo estimado y recursos didácticos

1 ½ hora	Introducción y discusión
2 horas (+ o -)	Trabajo en equipo

Introducción y discusión

Es más importante que los participantes comprendan el concepto del fenómeno de confusión y sepan identificar los factores potenciales de confusión, que el que entiendan plenamente cómo utilizar las fórmulas para determinar los efectos de dicho fenómeno.

Los facilitadores deben ayudar a los participantes en el cálculo de los riesgos relativos ajustados, de preferencia en la computadora.

Se recomienda que al menos un facilitador tenga un buen manejo del programa Epi Info, especialmente de Epi table. Sólo debería utilizarse la versión 6.04a o una versión posterior, puesto que las versiones anteriores tienen problemas en los cálculos de Epi table. El “Computer Companion” de la ISS (Vol. II, parte 2 de la serie de capacitación en ISS, Harare, Unidad de ISS de la OMS / AFRO, 1996) debe estar disponible durante el curso para utilizarse como referencia.

Módulo 27

PREPARACIÓN DEL ANÁLISIS ESTATÍSTICO: Medidas de dispersión, distribución normal y variaciones de la muestra

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 27 PREPARACIÓN DEL ANÁLISIS ESTADÍSTICO

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Explicar* qué se entiende por rango, percentil, desviación estándar, distribución normal, error estándar e intervalo de confianza al 95%.
2. *Calcular* rangos, desviaciones estándar, errores estándar e intervalos de confianza al 95% para sus propios datos cuando así convenga.
3. *Interpretar* los resultados de estos cálculos.

I. Introducción

II. Medidas de dispersión

III. Distribución normal

IV. Relación de la muestra con toda la población de estudio

V. Cómo determinar la medida en que los hallazgos en la muestra son representativos de la población.

I. INTRODUCTION

Como vimos en el *módulo 22*, la media, la mediana y la moda son medidas de la tendencia central de una variable, pero no provén ninguna información sobre la variación o la distribución de las mediciones. En este módulo se describen algunas medidas comunes de la variación (o variabilidad), que en los libros de estadística suelen llamarse medidas de dispersión. Además, se presentan los conceptos de distribución normal, error estándar e intervalo de confianza. Vamos a necesitar estos conceptos más adelante, en los *módulos 28-31*, cuando hagamos pruebas estadísticas.

II. MEDIDAS DE DISPERSIÓN

1. Rango

El *recorrido* de una serie de mediciones es la diferencia entre los valores máximo y mínimo.

Por ejemplo, si el peso de 7 mujeres embarazadas fue 40, 41, 42, 43, 44, 47 y 72 kg, el recorrido sería $72 - 40 = 32$ kg.

Si bien su cálculo es muy simple, el recorrido no nos dice nada de la distribución de los valores entre los dos extremos.

Si el peso de otras 7 mujeres embarazadas fuera 40, 46, 50, 55, 60, 65 y 72 kg, el recorrido también sería $72 - 40 = 32$ kg, aun cuando los valores son muy diferentes de los del ejemplo anterior.

2. Percentiles

Una segunda manera de describir la variación o dispersión de una serie de mediciones es dividiendo la distribución en percentiles. De hecho, el concepto de percentiles es simplemente una extensión del concepto de mediana, también llamada percentil 50.

Los *percentiles* son puntos que dividen todas las medidas en 100 partes iguales.

El tercer percentil (P3) es el valor por debajo del cual están 3% de las mediciones.

El percentil 50 (P50), o la mediana, es el valor por debajo del cual están 50% de las mediciones.

Para determinar los percentiles, las observaciones deben primero listarse de la más baja a la más alta, tal como se hace para obtener la mediana.

El concepto de percentiles es empleado por los nutricionistas para elaborar curvas de crecimiento estándar para determinados países, a partir de una población representativa de niños cuyo peso y estatura se miden de acuerdo con su edad en meses.

3. Desviación estándar

Para determinar qué tanto difieren nuestras mediciones del valor de la media, hay una medida, llamada la desviación estándar, que empleamos cuando aplicamos pruebas estadísticas.

La *desviación estándar* es una medida que describe cuánto difieren de la media, en promedio, las mediciones individuales.

Para obtener la desviación estándar de una serie de mediciones deben seguirse los pasos descritos a continuación:

1. Calcule la media de todas las mediciones.
2. Calcule la diferencia que hay entre cada medición individual y la media.
3. Eleve al cuadrado cada una de estas diferencias.
4. Obtenga la suma de todas las diferencias elevadas al cuadrado.
5. Divida esta suma entre el número de mediciones, menos uno.
6. Finalmente (puesto que las diferencias de la media se elevaron al cuadrado), saque la raíz cuadrada del valor obtenido (para volver a la misma unidad de medición).

Ejemplo 1:

Se pesó a 11 niños de 3 años de edad y se obtuvieron los siguientes valores:

13, 14, 14, 15, 16, 16, 16, 17, 17, 18 y 20 kg

El número de mediciones (n) es 11.

Para calcular la desviación estándar:

- (1) Calculamos primero la media: el valor de la media es 16 kg.

(2) Después, calculamos la distancia que hay entre cada medición y la media (16-13 = 3; 16-14 = 2; etc.): 3, 2, 2, 1, 0, 0, 0, 1, 1, 2, 4

(3) Elevamos al cuadrado estos valores (3 x 3 = 9; 2 x 2 = 4; etc.): 9, 4, 4, 1, 0, 0, 0, 1, 1, 4, 16

(4) La suma de estas diferencias elevadas al cuadrado (9 + 4 + ...) es 40.

(5) Esta suma se divide entre el número de mediciones (n) menos uno (n - 1 = 10):

$$\frac{40}{10} = 4$$

(6) Finalmente, sacamos la raíz cuadrada para obtener la desviación estándar de la media:

$$\sqrt{4} = 2\text{kg}$$

Una desviación estándar grande muestra que hay una gran dispersión de los valores alrededor de la media, mientras que una desviación estándar pequeña muestra que los valores individuales se concentran alrededor de la media, con poca variación entre ellos.

De hecho, si los pesos del *ejemplo 1* fueron

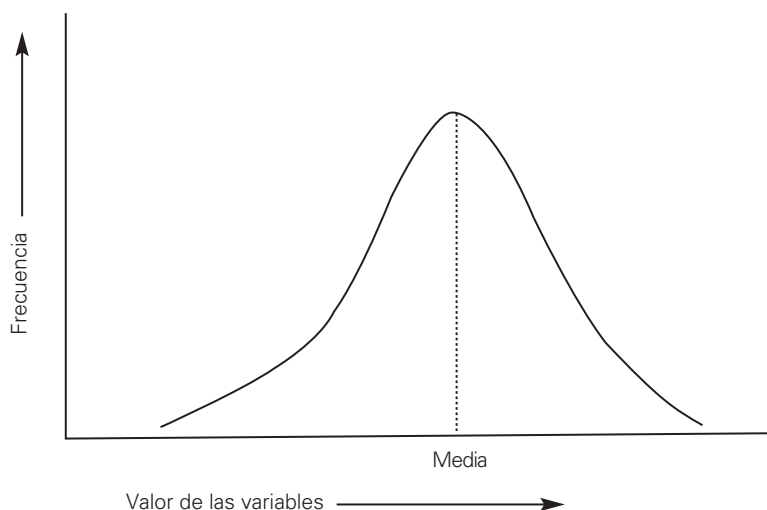
10, 11, 12, 14, 16, 16, 16, 18, 20, 21 y 22 kg,

La media habría seguido siendo 16 kg, pero la desviación estándar habría sido 4 kg, valor que indica una mucho mayor variación de las observaciones.

La fórmula que acabamos de emplear para calcular la desviación estándar se expone en el *anexo 27.1*. Otra manera de calcular una desviación estándar es empleando la fórmula expuesta en el *anexo 27.2*. Por fortuna, muchas calculadoras de bolsillo pueden hacer estos cálculos por nosotros, pero aun así, es importante entender lo que significa.

III. DISTRIBUCIÓN NORMAL

Muchas variables –tales como la estatura, el peso y la edad– tienen una *distribución normal*. Esta distribución da una curva en forma de campana con la mayoría de los valores agrupados cerca de la media y unos cuantos cerca de los extremos. La distribución normal es simétrica en torno a la media. La media, la mediana y la moda de una distribución normal tienen el mismo valor.

Ilustración 27.1 Curva de distribución normal

En la *ilustración 27.2* se muestra la curva normal dibujada sobre un histograma de la estatura de las mujeres embarazadas que asisten a una clínica de atención prenatal.

Una característica importante de una variable de distribución normal es que el 95% de las mediciones tienen valores que están aproximadamente dentro de 2 desviaciones estándar (DS) de la media. Esto se muestra en la siguiente ilustración.

Ejemplo 2:

Si la estatura media de un grupo de 120 mujeres es de 158 cm y la desviación estándar es de 3 cm, esto significa que alrededor de **95% de** las mujeres miden entre 152 y 164 cm (suponiendo que la distribución de las estaturas es normal). En otras palabras, 2.5% de las mujeres (porcentaje que corresponde a 3 mujeres) miden menos de 152 cm y 2,5% (o 3 mujeres) miden más de 164 cm.

Muchas pruebas estadísticas requieren que las variables tengan una distribución normal. Por lo tanto, es importante examinar la frecuencia de la distribución de las variables para determinar cuáles no tienen una distribución más o menos normal. Fuera de los ejemplos expuestos en los módulos vistos hasta aquí, las hay que no tienen una distribución normal (*ejemplo 4 del módulo 22*). Para evaluar si una variable tiene una distribución normal, utilice su representación gráfica como un histograma o un gráfico de línea (véase el módulo 22).

Ilustración 27.2 La estatura de mujeres embarazadas que asisten a una clínica de atención prenatal

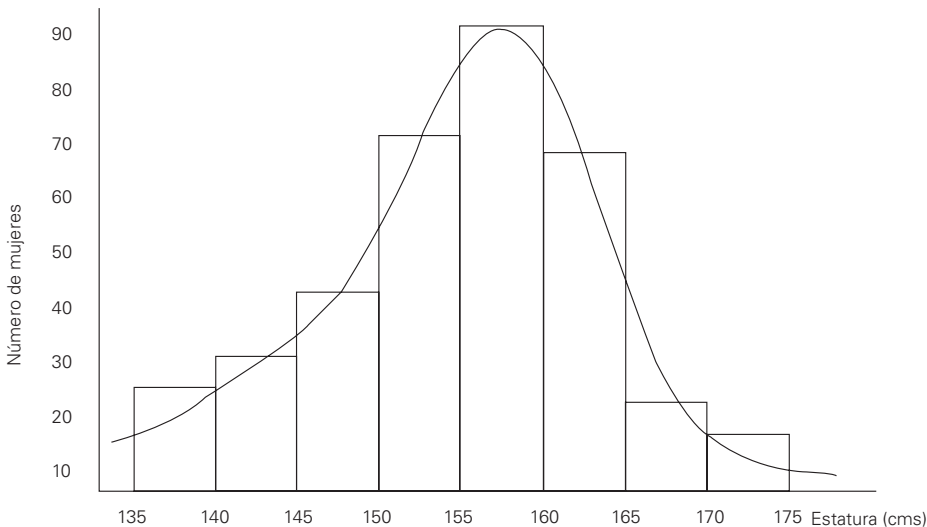
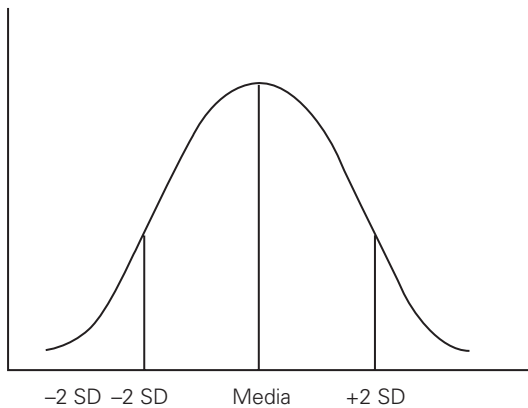


Ilustración 27.3 Una variable de distribución normal



IV. RELACIÓN DE LA MUESTRA CON TODA LA POBLACIÓN DE ESTUDIO

Cuando se emprende una investigación, por lo general es necesario establecer una muestra de la(s) población(es) de estudio. Enseguida se debe describir cada población a partir de la información recogida sobre la muestra. En otras palabras, se trata de generalizar los hallazgos de la muestra a una población de estudio

más amplia. Esto sólo puede hacerse si la muestra se seleccionó de tal manera que pueda considerarse representativa de la población total del estudio.

Cualquier valor de una variable obtenida para la muestra (por ejemplo, la media de la muestra) puede considerarse como una estimación del valor correspondiente para la población de estudio (la media de la población).

Por ejemplo, si 158 cm es la estatura media calculada para una muestra de 120 mujeres de determinado poblado, se espera que ésta sea una buena aproximación de la estatura media de toda la población de mujeres de ese poblado.

Sin embargo, no es probable que la media de la muestra sea idéntica a la media de la población.

Si extrae otra muestra de 120 madres, podría encontrar una media de 157 cm, que no es idéntica a la media de la primera muestra. Es probable que ésta también difiera de la estatura media real de la población total de la cual se ha tomado la muestra.

Este fenómeno se denomina *variación (variabilidad) debida al muestreo*.

La *ilustración 27.4* muestra una curva de distribución de frecuencias de una población con las curvas de dos muestras diferentes en su interior.

Note que cualquier muestra representativa va a tener una curva de distribución de forma semejante a la curva de la población, pero puede caer en cualquier punto de la curva de la población.

V. CÓMO DETERMINAR LA MEDIDA EN QUE LOS HALLAZGOS DE LA MUESTRA SON REPRESENTATIVOS DE TODA LA POBLACIÓN

Para saber en qué medida un valor de la muestra se desvía del de la población, se calcula el rango o un intervalo en torno al valor de la muestra, que muy probablemente va a contener el valor de la población.

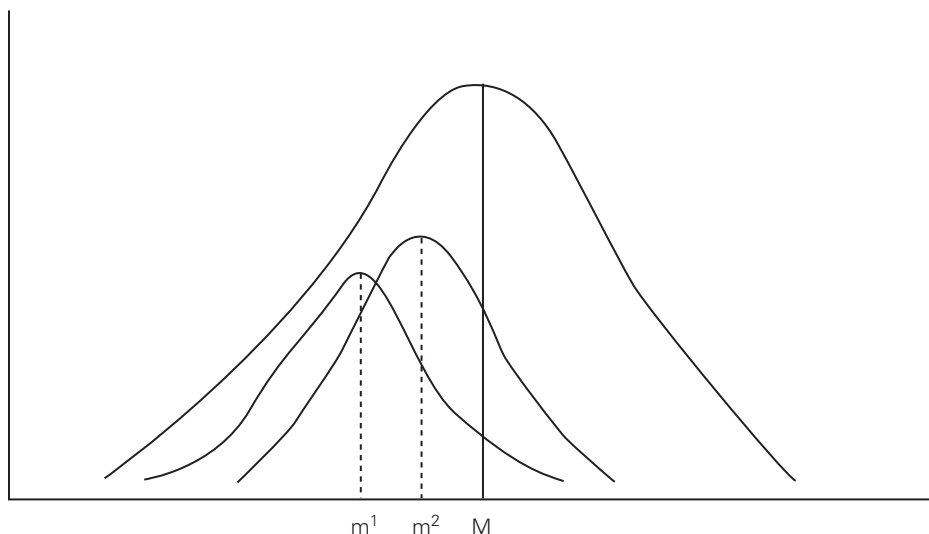
Este recorrido se llama *intervalo de confianza*.

Un intervalo de confianza es el intervalo o recorrido (amplitud) de los valores que tienen una mayor probabilidad de contener el valor real de la población. Los límites inferior y superior de este intervalo se denominan *límites de confianza*.

Por ejemplo:

El intervalo de confianza al 95% de 152 a 164 cm para la estatura media de una población de mujeres significa que usted tiene 95% de seguridad de que la me-

Ilustración 27.4 Curva de distribución de frecuencias y las curvas de dos muestras para una variable de distribución normal



dia real de la población, que usted no puede conocer con exactitud a menos que midiera la estatura de todas las mujeres, se encuentra entre 152 y 164 cm. (152 cm es el límite de confianza inferior, 164 cm es el límite de confianza superior).

El cálculo de un intervalo de confianza considera el *error estándar*. El error estándar es una estimación del grado en que la media de la muestra difiere de la media de la población. Éste se calcula sobre la base de la desviación estándar.

Ahora vamos a comentar cómo calcular:

- el error estándar y el intervalo de confianza al 95% para la media (para datos numéricos)
- el error estándar y el intervalo de confianza al 95% para un porcentaje (para datos categóricos)

1. Cómo calcular el error estándar y el intervalo de confianza al 95% para una media

Al tratar con *datos numéricos* usted quizás quiera saber en qué medida la media de la muestra difiere de la media de la población.

El error estándar de la media se calcula dividiendo la desviación estándar entre la raíz cuadrada del tamaño de la muestra:

$$EE = \frac{\text{desviación estándar}}{\sqrt{\text{tamaño de la muestra}}} \quad \text{o} \quad \frac{DE}{\sqrt{n}}$$

Para una variable de distribución normal, puede suponerse que aproximadamente 95% de todas las posibles medias de la muestra se encuentran dentro de dos errores estándar de la media de la población. En otras palabras, podemos estar seguros al 95% de que la media de la población, –de la que queremos tener la mejor estimación posible–, se encuentra dentro de dos errores estándar de la media de la muestra.

Cuando las variables se describen estadísticamente, por lo general se presenta la media de la muestra que se calculó, más/menos dos errores estándar. A esto se le llama *intervalo de confianza al 95%*. Esto significa que usted está aproximadamente 95% seguro de que la media real de la población está dentro de este intervalo.

En el *ejemplo 1*, se tomó el peso de una muestra aleatoria de 11 niños de 3 años en un determinado poblado. La media de la muestra fue de 16 kg y la desviación estándar de la muestra fue de 2 kg.

El error estándar es:

$$\frac{2}{\sqrt{11}} = 0,6 \text{ KG}$$

El intervalo de confianza al 95% es:

$$16 \pm (2 \times 0.6) = 14.8 \text{ a } 17.2 \text{ kg}$$

Esto significa que estamos aproximadamente 95% seguros de que el peso medio de todos los niños de tres años de edad en ese poblado está entre 14,8 y 17,2 kg.

Cuanto mayor es el tamaño de la muestra, menor es el error estándar y más estrecho es el intervalo de confianza (o el margen de error; véase en el módulo 11 “el tamaño de la muestra”). Así, la ventaja de una muestra bastante amplia es que su media va a ser una mejor estimación respecto de la población. Sin embargo, a partir de un cierto punto, aumentar el tamaño de la muestra implica una fuerte inversión de tiempo y dinero, mientras que el intervalo de confianza disminuye sólo ligeramente (véase el anexo 11.3).

En el ejemplo anterior, un incremento del tamaño de la muestra aumenta sin duda la fiabilidad del cálculo. Con un tamaño de la muestra de 20 (en lugar de 11), el error estándar es:

$$EE \frac{2}{\sqrt{20}} = 0,45 \text{ kg,}$$

Y el intervalo de confianza al 95% para la media del peso es de 15.1 a 16.9 kg.

2. Cómo calcular el error estándar y el intervalo de confianza al 95% para un porcentaje

En la sección anterior, calculamos el error estándar y el intervalo de confianza al 95% para la media de una muestra, a partir de *datos numéricos*. Ahora vamos a hacer lo mismo para un porcentaje que se calculó a partir de *datos categóricos*.

Ejemplo 3:

En una muestra de 120 pacientes con tuberculosis, extraída de la población total de pacientes con TB del país, se encontró que 28 de ellos (el 23,3%) no siguieron su tratamiento ambulatorio. El cumplimiento de los otros 92 (o 76,7%) fue satisfactorio. Queremos calcular ahora el error estándar del porcentaje de incumplimiento (23,3%). Esto se hace de la siguiente manera.

Si p representa uno de los porcentajes (23,3%) y $100 - p$ representa el otro (76,7%), entonces el error estándar del porcentaje se obtiene multiplicándolos, dividiendo el resultado entre el número de la muestra y sacándole la raíz cuadrada.

La fórmula para obtener el error estándar de un porcentaje es:

$$EE \sqrt{\frac{p(100 - p)}{n}}$$

En el *ejemplo* esto es:

$$\sqrt{\frac{23,3 \times 76,7}{120}} = 3,9$$

Ahora queremos calcular también el intervalo de confianza para el porcentaje de pacientes incumplidos de todo el país.

El intervalo de confianza al 95% es

$23,3\% \pm (2 \times 3,9)$ lo que está entre $15,5\%$ y $31,1\%$.

Esto significa que tenemos confianza al 95% en que de toda la población de los pacientes de TB del país donde se tomó la muestra de 120, de $15,5\%$ a $31,1\%$ no siguen su tratamiento ambulatorio.

Ahora bien, en vez de porcentajes podemos usar proporciones, que pueden tomar cualquier valor entre 0 y 1. La fórmula para obtener el error estándar es, entonces:

$$\sqrt{\frac{p(1-p)}{n}}$$

En el *ejemplo* esto es:

$$EE = \sqrt{\frac{0,233 \times 0,767}{120}} = 0,039$$

El intervalo de confianza al 95% está entre 0,155 y 0,311, o entre $15,5\%$ y $31,1\%$.

TRABAJO EN EQUIPO

1. Calcule el rango, la desviación estándar y el intervalo de confianza al 95% para sus series de datos numéricos más importantes.

Interprete los resultados de estos cálculos.

2. Calcule el intervalo de confianza al 95% de los porcentajes de sus series de datos categóricos más importantes. Este mismo cálculo puede hacerse para los datos numéricos, si se resumen en categorías.

Interprete los resultados de los cálculos.

3. Guarde los resultados de su trabajo en equipo. Los va a necesitar cuando haga las pruebas estadísticas.

Anexo 27.1 Método extendido para calcular la desviación estándar

x_i = las observaciones (el peso de los niños de tres años, véase el ejemplo 1)

$$\bar{x} = \text{media} = \frac{\sum x_i}{n}$$

$$x_i - \bar{x} = \text{diferencia} = d_i$$

La media $\bar{x} = 176/11 = 16$ kg

x_i	$\bar{x} - x_i$	d_i	d_i^2
13	$16 - 13$	+3	9
14	$16 - 14$	+2	4
14	$16 - 14$	+2	4
15	$16 - 15$	+1	1
16	$16 - 16$	0	0
16	$16 - 16$	0	0
16	$16 - 16$	0	0
17	$16 - 17$	-1	1
17	$16 - 17$	-1	1
18	$16 - 18$	-2	4
20	$16 - 20$	-4	<u>16</u>
			40

$$\sum d_i^2 = 40$$

La desviación estándar es:

$$\sqrt{\frac{\sum d_i^2}{n-1}} = \sqrt{40/10} = 2 \text{ kg}$$

Anexo 27.2 Método breve para calcular la desviación estándar

Desviación estándar (DE) =

$$\sqrt{\frac{\sum x^2 - (\sum x)^2/n}{n-1}}$$

donde x = cada valor
 x^2 = el cuadrado de cada valor
 (sigma) Σ “la suma de”
 n = el número de observaciones

Si aplicamos esta fórmula en el *ejemplo 1* (el peso de 11 niños de tres años), va usted a descubrir que no es tan difícil como parece al principio. Los valores x son:

13, 14, 14, 15, 16, 16, 16, 17, 17, 18, 20 kg y $n = 11$

Si elevamos cada valor al cuadrado, obtenemos:

169, 196, 196, 225, 256, 256, 256, 289, 289, 324, 400

La suma de todos estos cuadrados ($\sum x^2$) es 2856

La suma de todos los valores ($\sum x$) es 176

Por lo tanto $(\sum x)^2 = 176^2 = 30,976$

$$\text{y } (\sum x)^2 = \frac{30.976}{11} = 2.816$$

Desviación estándar =

$$\sqrt{\frac{(2856 - 2816)}{10}} = \sqrt{4} = 2 \text{ kg}$$

Nota:

Un método más rápido para calcular la desviación estándar es usando la función automática incorporada en muchas calculadoras de bolsillo.

Módulo 27 PREPARACIÓN DEL ANÁLISIS ESTADÍSTICO

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
3 horas +	Trabajo en equipo

Introducción y discusión

- Éste es el primer módulo en el que se introducen conceptos estadísticos, que para algunos participantes –si no es que para todos– son nuevos, dependiendo de su nivel académico. Por lo tanto, se le recomienda irse despacio, ser cuidadoso con las fórmulas y emplear ejemplos sencillos.
- Es crucial que el concepto de *variación de la muestra* (parte IV) les quede claro a todos, para entender cómo funcionan las pruebas de significación (*módulo 28*).
- Haga el vínculo con el *módulo 11* y muestre el *anexo 11.3* para visualizar los rendimientos decrecientes de muestras cada vez mayores.
- Si los participantes tienen poca experiencia con las estadísticas puede dejar fuera las secciones 1 y 2 de la Parte V de su presentación. Sin embargo, si uno de los grupos está haciendo un estudio descriptivo y no ha hecho ninguna comparación entre los grupos, quizás necesita calcular el error estándar de la media (sección 1) o el error estándar de un porcentaje (sección 2), en lugar de hacer pruebas estadísticas. En ese caso, conserve esta sección cuando haga su presentación.

Trabajo en equipo

- Si algunos de los cálculos no son adecuados para ciertos grupos, deben hacer por lo menos un ejercicio de cada uno empleando datos de la muestra para tener la experiencia.
- Considere que no hay necesidad de calcular los errores estándar para las diferencias o para los porcentajes, si es que se van a hacer pruebas estadísticas para determinar las diferencias entre los grupos o para medir las asociaciones entre las variables. (*Véanse los módulos 28 a 31*).

Módulo 28

SELECCIÓN DE LAS PRUEBAS DE SIGNIFICACIÓN

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 28 SELECCIÓN DE UNA PRUEBA DE SIGNIFICACIÓN

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Explicar* qué es una prueba de significación y cuál es su propósito.
2. *Usar* las tablas para seleccionar pruebas de significación para diferentes conjuntos de datos.
3. *Elegir* pruebas de significación apropiadas para sus datos.

I. Introducción

II. Pruebas de significación

III. Cómo funcionan las pruebas de significación

IV. Selección de una prueba de significación

I. INTRODUCCIÓN

A lo largo de este curso se ha subrayado que el análisis y la interpretación de los resultados del estudio deben estar relacionados con los objetivos. El *módulo 24* explica cómo construir tabulaciones cruzadas adecuadas para los objetivos de la investigación. Quizás hayamos encontrado algunos resultados interesantes. *Por ejemplo*, en un estudio sobre el consumo de alcohol, descubrimos que 30% de los hombres y 20% de las mujeres incluidos en la muestra toman más de tres vasos de alcohol al día. ¿Cómo debemos interpretar este resultado?

- La diferencia observada de 10% podría ser una *diferencia verdadera*, que existe también en la población total de la que se extrajo la muestra.
- La diferencia también puede *deberse a la casualidad*; en realidad puede no haber ninguna diferencia entre los hombres y las mujeres, y simplemente la muestra de los hombres haber diferido de la muestra de mujeres. También puede deberse a una variación de la muestra.
- Una tercera posibilidad es que la diferencia de 10% se deba a defectos (o *sesgos*) en el diseño del estudio. Por ejemplo, que hubiéramos entrevistado sólo a hombres u omitido hacer una pre-prueba. Muchas veces el alcoholismo es un tema muy delicado para la mujer, y es necesario, por lo tanto, diseñar un abordaje flexible. Con un diseño de estudio apropiado no se habría encontrado semejante diferencia.

Si estamos seguros de que la diferencia entre dos grupos no se explica por un sesgo, sólo podemos concluir que se trata de una diferencia verdadera si descartamos *la casualidad* (variación de la muestra, véase el módulo 27) mediante una prueba de significación.

Una PRUEBA DE SIGNIFICACIÓN estima la probabilidad de que una observación resultante del estudio (por ejemplo, una diferencia entre dos grupos) se deba a la casualidad.

En otras palabras, una prueba de significación se utiliza para averiguar si el resultado de un estudio –observado en una muestra–, puede considerarse como un resultado que existe efectivamente en toda la población.

II. PRUEBAS DE SIGNIFICACIÓN

Cada conjunto de datos requiere de una prueba de significación particular. A lo largo de este módulo, y los *módulos 29-31*, distinguimos dos grandes conjuntos de datos.

- Dos (o más) *grupos* que se comparan para detectar *diferencias*. (Por ejemplo, de hombres y de mujeres, para detectar diferencias en el consumo del alcohol).
- Dos (o más) *variables* que se miden para determinar si hay una *asociación* entre ellas. (Por ejemplo, entre el consumo de alcohol y el ingreso).

A fin de ayudarle a elegir la prueba correcta, vamos a presentar un diagrama de flujo y matrices para diferentes conjuntos de datos. En los siguientes módulos (29-31) explicamos los pormenores de algunas de las pruebas de significación más usuales. Pero ahora vamos a comentar su funcionamiento.

III. CÓMO FUNCIONAN LAS PRUEBAS DE SIGNIFICACIÓN

El razonamiento detrás de toda prueba de significación es el mismo, ya sea que el investigador esté comparando dos grupos en busca de sus diferencias o que esté midiendo dos variables para detectar asociaciones.

Vamos a concentrarnos, en primer lugar, en la *comparación de grupos*.

- Suponga que observó una diferencia entre dos grupos de su muestra.
- Desea saber si esta diferencia entre los grupos representa una diferencia real en la población total del estudio –de la que se extrajo la muestra–, o si *ocurrió por mera casualidad* (debido a la variación de la muestra).
- Para averiguarlo, determine qué tan probable es que esta diferencia haya podido ocurrir por casualidad; si en la población total no hay efectivamente tal diferencia entre los dos grupos.

Nunca se está 100% seguro de la realidad de una diferencia. En general nos sentimos satisfechos si podemos estar 99% o 95% seguros de que el resultado sea cierto para la población total del estudio. Si estamos 95% seguros, la probabilidad de que la diferencia observada haya ocurrido por casualidad es inferior a 5%. Así, seleccionamos un nivel comúnmente aceptado para la posibilidad de que nuestra conclusión haya *ocurrido por casualidad*, 0,10 (10%), 0,05 (5%), 0,01 (1%) o incluso 0,001 (0,1%). Esto se conoce como el nivel de significación elegido (también denominado α , o nivel alfa). Puede emplearse, asimismo, la expresión *nivel de confianza*.

Si usted quiere estar tan seguro como sea posible de que la diferencia que observó es cierta, va a elegir el nivel de significación 0,01 o 0,001. Un ensayo sobre los nuevos fármacos seleccionaría estos niveles. En muchos estudios de gestión o de la conducta no podemos estar tan seguros y aceptamos un nivel de significación de 0,05. Los valores 0,01 y 0,05 son los más empleados en los estudios científicos.

En cualquier estudio que busque diferencias entre grupos o asociaciones entre variables, la *probabilidad* (p) de observar un determinado resultado por casualidad debe calcularse mediante pruebas estadísticas.

Esta *probabilidad* de que un resultado observado se deba a la casualidad se expresa normalmente como un *valor-p*.

En el *estudio del alcohol*, el valor de p –calculado para determinar si la diferencia observada entre hombres y mujeres en la conducta de consumo de alcohol se debió a la casualidad– fue 0.009.

El nivel de significación elegido fue 0.01 (1%), que es inferior a 0,009. Podemos, por lo tanto, estar más de 99% seguros de que los hombres beben más que las mujeres en la población de estudio seleccionada. Podemos decir, entonces, que este resultado es estadísticamente significativo en el nivel 0.01.

Si el valor de p hubiera sido superior a 0.01 (por ejemplo, 0.03), el resultado *no* habría sido estadísticamente significativo en el nivel 0.01.

El mismo razonamiento se aplica a las *asociaciones entre las variables*.

Supongamos que categorizamos la conducta de consumo de alcohol en cuatro grupos: nunca, moderada, grave, excesiva. Los grupos de nivel de ingresos se categorizaron como bajo, medio y alto. Observamos la siguiente tendencia: la conducta de consumo de alcohol es más grave o excesiva en los grupos de menor ingreso. Ahora debe calcularse la probabilidad de que esta asociación entre el consumo de alcohol y el ingreso se deba a la casualidad. El valor de p calculado es de 0,07. Habíamos elegido un nivel de significación del 0,05%. Como nuestro valor p es superior a 0,05, este resultado no es estadísticamente significativo en el nivel de 0,05, y no podemos estar 95% seguros de que la asociación entre el consumo de alcohol y el nivel de ingresos sea real.

Nota:

En términos estadísticos, el supuesto de que en la población total de estudio no hay una diferencia real entre los grupos –o de que no existe una asociación real entre las variables–, se llama la *hipótesis nula* (H_0). La *hipótesis alternativa* (H_a) es que sí hay una diferencia entre los grupos o que sí existe una asociación real entre las variables.

Los siguientes son algunos ejemplos de hipótesis nula:

- No hay diferencia en la incidencia de sarampión entre los niños vacunados y los que no están vacunados

- Los hombres no beben más alcohol que las mujeres
- No hay asociación entre el ingreso de las familias y la desnutrición de los niños

Nota:

Si el resultado es estadísticamente significativo, rechazamos la *hipótesis nula* (H_0) y aceptamos la *hipótesis alternativa* (H_a) de que sí hay una diferencia real entre dos grupos o una asociación real entre dos variables.

Los siguientes son algunos ejemplos de hipótesis alternativa:

- Sí hay una diferencia en la incidencia de sarampión entre los niños vacunados y los que no están vacunados
- Los hombres beben más alcohol que las mujeres
- Sí hay una asociación entre el ingreso de las familias y la desnutrición de los niños

Tenga presente que “estadísticamente significativo” no quiere decir que una diferencia o una asociación tenga *importancia práctica*. La diferencia más pequeña e irrelevante va a resultar estadísticamente significativa si la muestra es lo suficientemente grande. Por otra parte, una diferencia importante puede no alcanzar significación estadística si la muestra es demasiado pequeña.

Es importante que el investigador establezca si los resultados que son estadísticamente significativos lo son también prácticamente. La significación práctica implica que el problema justifica que se emprenda algún tipo de acción para subsanarlo. Por ejemplo, si la diferencia en la incidencia de sarampión entre los 500 niños vacunados y los 500 sin vacunar es de 35%, la vacunación debe promoverse de una manera más activa. Sin embargo, si la diferencia es de sólo 2%, puede no justificarse ninguna acción adicional.

IV. SELECCIÓN DE UNA PRUEBA DE SIGNIFICACIÓN

Dependiendo del propósito del estudio y el tipo de datos recogidos, debe usted seleccionar una prueba de significación apropiada.

Nota:

Antes de aplicar cualquier prueba de significación, establezca la hipótesis nula en relación con los datos para los que se está aplicando la prueba. Esto le va a permitir interpretar los resultados.

La siguiente sección explica cómo seleccionar una prueba de significación apropiada para determinar las diferencias entre los grupos o las asociaciones entre las variables.

1. Determinación de las diferencias significativas entre los grupos (utilizando la ilustración 28.1)

Hay varias cuestiones que debe usted considerar al decidir qué prueba usar para determinar si las diferencias entre dos grupos son estadísticamente significativas. Primero vea si tiene observaciones apareadas¹ o sin aparear. (De ser necesario, consulte los *módulos 9 y 26*). Dentro de cada una de estas categorías, determine si los datos son categóricos (nominales y ordinales) o numéricos. (Para revisar las definiciones consulte los *módulos 8 y 22*.)

Para los datos nominales (pareados o sin aparear) la prueba de significación que debe emplearse depende de si la muestra es pequeña o grande. No hay un criterio claro sobre lo que debe considerarse “pequeño” o “grande”. Sin embargo, en el caso de las observaciones sin aparear es mejor utilizar la *prueba exacta de Fisher*, en vez de la *prueba de chi cuadrado* (χ^2), si la muestra total es inferior a 40 o si cualquier celda de la tabla –que debe construirse–, tiene un número previsto menor de 5. En estos módulos de capacitación sólo consideramos las pruebas utilizadas con muestras mayores. La *prueba de chi cuadrado* (χ^2) se trata en el *módulo 29* y la *prueba de chi cuadrado de McNemar*, en el *módulo 30*.

Ejemplo de una muestra sin pares:

En un estudio de la eficacia de la vacunación contra el sarampión, el equipo de investigación estudia un grupo de 100 pacientes de entre 1 y 5 años de edad que ingresan en una clínica con diagnóstico de sarampión, y un grupo de 100 niños en el mismo rango de edad cuyo ingreso se debe a otras enfermedades. Al comparar el registro de vacunas de los dos grupos, se encontró que la tasa de vacunación entre los niños con sarampión era inferior a la de los pacientes que no tenían sarampión. Se utilizó la prueba de chi cuadrado (χ^2) para comprobar si esta diferencia era estadísticamente significativa.

Utilizando *la ilustración 28.1*, éstos son los pasos que deben seguirse para determinar qué prueba emplear.

¹ Se habla de observaciones apareadas cuando las observaciones individuales de un conjunto de datos (por ejemplo, los casos) se corresponden con las observaciones individuales de otro conjunto de datos (por ejemplo, los controles), siempre y cuando los participantes en el estudio de uno y otro grupo provengan del mismo lugar o tengan la misma edad.

1. *Objetivo*: estudiar la eficacia de las vacunas contra el sarampión entre los niños de 1 a 5 años de edad.
2. *Hipótesis*:
 H_0 : no hay ninguna diferencia en la tasa de vacunación entre los pacientes con sarampión y los que no tienen sarampión.
 H_a : sí hay una diferencia en la tasa de vacunación entre los pacientes con sarampión y los que no tienen sarampión.
3. *Conjunto de datos*: comparación de grupos (tasa de vacunación).
4. *Tipo de observación*: impar.
5. *Tipo de variable*: categórica (nominal).
6. *Muestra*: grande.
7. *Prueba*: prueba χ^2 .

Para datos *ORDINALES* la prueba de significación que debe usarse depende de si se va a comparar sólo dos grupos o más.²

Para datos *NUMÉRICOS*, al igual que para los datos ordinales, la selección de una prueba de significación apropiada depende de si se va a comparar dos grupos o más de dos grupos.

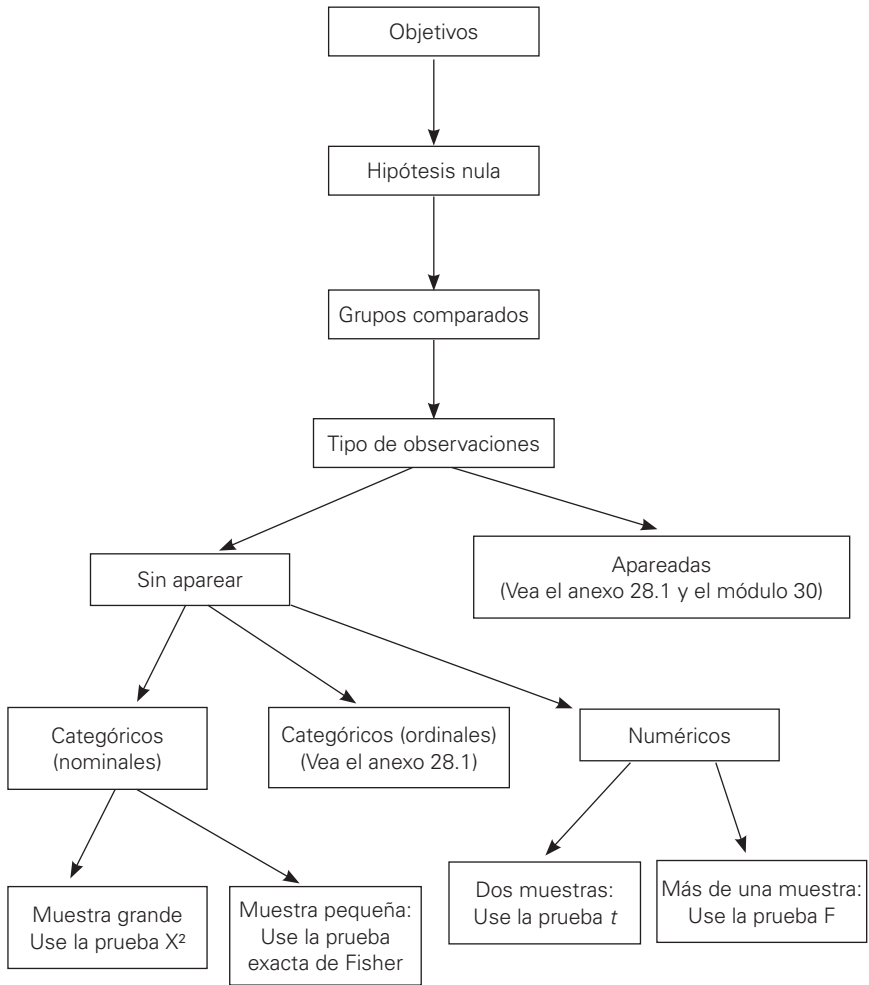
En el *módulo 29* vamos a comentar la forma de realizar e interpretar una *prueba t* si usted va a comparar sólo dos grupos. Si debe comparar más de dos grupos, consulte un libro de texto de estadística. (Véanse la ilustración 28,1 y la tabla 28.1).

Ejemplo de una muestra no pareada en dos grupos:

En un estudio de nutrición, se midió el peso de 142 niños de cinco años de edad que vivían en zonas rurales y de 171 niños de esa misma edad que vivían en zonas urbanas. Se calculó el peso medio de cada una de las muestras y se comparó utilizando la *prueba t*, para determinar si había una diferencia.

2 Las pruebas que se utilizan para comparar dos grupos se basan en el ordenamiento de los datos: la *prueba de Wilcoxon para dos muestras*, que da resultados equivalentes a los de la *prueba de la U de Mann-Whitney*, para las observaciones sin parear, y la *prueba de Wilcoxon de los rangos con signos* para las observaciones pareadas. No vamos a tratar estas pruebas en nuestros módulos de capacitación, pero si usted desea utilizarlas, por favor consulte un libro de texto de estadística. (Véanse las referencias propuestas en este módulo; el capítulo 10 del libro *Statistics at Square*, de Dougal Swinscow, es muy claro y fácil de entender.)

Ilustración 28.1 Diagrama de flujo para seleccionar la prueba de significación para determinar las diferencias entre dos grupos



Utilizando la *ilustración 28.1*, éstos son los pasos que deben seguirse para determinar qué prueba emplear.

1. *Objetivo*: determinar las diferencias de peso entre los niños de cinco años de las zonas rurales y los de las zonas urbanas.
2. *Hipótesis*:

H_0 : no hay diferencia entre el peso medio de los niños de zonas urbanas y el de los niños de zonas rurales.

H_a : hay diferencia entre el peso medio de los niños de zonas urbanas y el de los niños de zonas rurales.

3. *Conjunto de datos*: comparación de grupos (media del peso).
4. *Tipo de observación*: impar.
5. *Tipo de variable*: numérica.
6. *Muestra*: dos muestras.
7. *Prueba*: prueba *t*.

Ejemplo de una muestra impar en más de dos grupos:

Se comparó el peso medio de los siguientes 4 grupos de niños de cinco años: niños que viven en zonas rurales, niños que viven en zonas urbanas, niñas que viven en zonas rurales y niñas que viven en zonas urbanas. En este caso, la *prueba F* fue la elección adecuada.

Ejemplo de un estudio en que los datos están ordenados:

En un estudio cuasi-experimental para investigar el efecto de una campaña de educación sanitaria sobre el conocimiento para manejar la diarrea en la comunidad, se seleccionan dos grupos de poblados. En el primer grupo de poblados sí se realizó la campaña, y en el segundo no. Durante el análisis, los poblados se clasificaron por su nivel de conocimiento sobre el tratamiento adecuado de la diarrea, del más alto al más bajo. Para determinar si había una diferencia significativa entre los dos grupos de poblados se hizo la *prueba de Wilcoxon para dos muestras*.

Ejemplo de una muestra apareada, en dos grupos:

Se compara el peso medio de las mujeres y los hombres adultos, controlando la estatura. Esto significa que por cada hombre de una cierta estatura, se seleccionó una mujer de la misma estatura, de manera que pudiera compararse el peso de cada par. En este caso, se utilizó la *prueba t para datos apareados*.

La tabla 28.1 resume las diferentes pruebas disponibles para determinar si las diferencias entre dos o más grupos de observaciones apareadas o sin aparear son significativas.

Tabla 28.1 Selección de una prueba de significación para determinar las diferencias entre grupos

	<i>Observaciones sin aparear</i>	<i>Observaciones apareadas</i>
<i>Datos nominales</i>		
Muestra pequeña	Prueba exacta de Fisher	Prueba con signos
Muestra grande	Prueba de chi cuadrado (χ^2) * (módulo 29)	Prueba de chi cuadrado (χ^2) de McNemar* (módulo 30)
<i>Datos ordinales</i>		
Dos grupos	Prueba de Wilcoxon para dos muestras o prueba de la U de Mann-Whitney	Prueba de los rangos con signo de Wilcoxon
Más de dos grupos	Análisis de la varianza de una vía de Kruskal-Wallis	Análisis de la varianza de dos vías de Friedman
<i>Datos numéricos</i>		
Dos grupos	Prueba t* (módulo 29)	Prueba t para datos apareados* (módulo 30)
Más de dos grupos	Prueba F	

* Las pruebas seguidas de un asterisco se comentan en los módulos señalados.

Fuente: Adaptación de Riegelman RK (1981).

2. Medición de las asociaciones entre variables (empleando la tabla 28.2)

Tabla 28.2 Selección de una prueba de significación para medir las asociaciones entre variables

<i>Datos nominales</i>	Prueba de chi cuadrado (χ^2), (si la muestra es lo suficientemente grande)* (módulos 29, 30)	Cálculo de la razón de proba- bilidades o cálculo del riesgo relativo* (módulo 25)
<i>Datos ordinales o datos numéricos cuando no se sospecha que pueda haber una relación lineal</i>	Cálculo del coeficiente de co- rrelación rho de Spearman o del coeficiente de correlación tau de Kendall	Significación del coeficiente de correlación rho de Spearman o del coeficiente de correlación tau de Kendall
<i>Datos numéricos cuando se sospecha que pueda haber una relación lineal</i>	Cálculo del coeficiente de correla- ción de Pearson (r)* (módulo 31)	Significación del coeficiente de correlación de Pearson (r)* (módulo 31)

* Las pruebas seguidas de un asterisco se comentan en los módulos señalados.

Fuente: Adaptación de Riegelman RK (1981).

Determine si sus datos son nominales, ordinales o numéricos. Si son numéricos, decida si hay la sospecha de una relación lineal. La expresión “relación lineal” para los datos numéricos significa que la asociación es tal que la variable dependiente cambia en una relación constante con la variable independiente, de tal manera que los puntos de un diagrama de dispersión, cuando se unen, se aproximan mejor por una línea recta.

Para los *datos NOMINALES*, el *riesgo relativo* es una medida útil de la asociación, que se aplica con frecuencia en los estudios de casos y controles y en los de cohorte. El *módulo 25* aborda el cálculo de la razón de probabilidades como estimación del riesgo relativo en los estudios de casos y controles.

Ejemplo:

En el estudio de casos y controles presentado en el *ejemplo 3* de dicho módulo, empleamos la razón de probabilidades para calcular que los recién nacidos con bajo peso tenían un riesgo de muerte neonatal 12 veces mayor que los recién nacidos con peso normal.

Para determinar la significación de los datos *ORDINALES*, puede calcular y probar el *coeficiente de correlación de Spearman (rho)* o el *tau de Kendall*. Si desea utilizarlos, consulte un libro de texto de estadística. (Vea usted las referencias que se hacen, por ejemplo, en Swinscow (1998), capítulo 11).

Para determinar la significación de los datos *NUMÉRICOS* cuando sospecha que hay una relación lineal, puede calcular y probar el *coeficiente de correlación de Pearson*. El *módulo 31* expone cómo hacerlo.

Ejemplo:

Puede usted examinar si el peso de los niños de cinco años de edad se asocia al ingreso familiar. Sospecha que podría haber una relación lineal entre las dos variables: “ingreso familiar” y “peso”, tal que el peso aumenta con el incremento del ingreso familiar.

Nota:

Si una asociación es estadísticamente significativa no implica necesariamente que haya una relación causal (véase el módulo 25). Sin embargo, a menudo suscita una mayor investigación para averiguar si hay o no una relación causal.

EJERCICIOS de selección de pruebas

Utilice la *ilustración 28.1* y las *tablas 28.1* y *28.2* para identificar las pruebas adecuadas para los siguientes estudios de investigación.

Ejercicio 1:

Se hace un estudio para comparar el efecto de un nuevo medicamento contra la hipertensión en la presión arterial diastólica de un grupo muestra con el efecto de un placebo en un grupo de control sin aparear (Riegelman 1981: 243).

Ejercicio 2:

Se hace un estudio para determinar si entre las mujeres embarazadas que viven en hogares donde no hay suministro de agua propio hay un riesgo significativamente mayor de muerte perinatal que entre las mujeres embarazadas en cuyo hogar hay un suministro de agua independiente. Si así fuera, el estudio va a medir qué tan fuerte es esta asociación.

Ejercicio 3:

Se emprende un estudio para determinar si hay una pérdida significativa de peso después de un año de tratamiento de la diabetes, y si el peso perdido se relaciona con el peso inicial. La siguiente tabla expone el peso inicial (x) y el peso después de un año de terapia (y) de 16 pacientes adultos recién diagnosticados de diabetes.

<i>Peso inicial (x) en libras</i>	<i>Peso después de un año (y) en libras</i>	<i>Peso inicial (x) en libras</i>	<i>Peso después de un año (y) en libras</i>
140	115	120	123
160	130	145	143
180	135	150	125
120	125	160	140
132	112	160	135
146	130	149	120
190	160	129	119
200	160	150	113

Ejercicio 4:

Estudios anteriores determinan que 30% de las parejas elegibles en un distrito de salud practican la planificación familiar. Después de un programa de educación masiva, los resultados indican que de 90 parejas elegibles seleccionadas al azar, 40 practican la planificación familiar. El funcionario responsable de la educación para la salud quiere saber si su programa ha afectado al grupo objetivo.

TRABAJO EN EQUIPO

Considerando los objetivos específicos de su estudio y la lista de variables, y empleando las tabulaciones cruzadas ya realizadas, identifique las pruebas de significación que tendrá que hacer con sus datos.

REFERENCIAS

- Altman DG (1991). *Practical statistics for medical research*, London: Chapman and Hall.
- Barker DJP (1982). *Practical epidemiology*, (3rd ed.), Edinburgh, UK: Churchill Livingstone.
- Castle WM and North PM (1995). *Statistics in Small Doses*, Edinburgh, UK: Churchill Livingstone.
- Bradford Hill A (1984). *A short textbook of medical statistics*, (11th ed.), London, UK: Hodder and Stoughton.
- Kelsey JL, Thompson WD and Evans AS (1986). *Methods in observational epidemiology*, Oxford, UK: Oxford University Press.
- Kidder LH and Judd CM (1986). *Research methods in social relations*, New York, USA: CBS College Publishing.
- Kleinbaum DG, Kupper LL, Morgenstern H (1982). *Epidemiologic research - principles and quantitative methods*, New York, USA: Van Nostrand Reinhold.
- Riegelman RF (1981). *Studying a study and testing a test*, Boston, MA, USA: Little Brown and Company.
- Schlesselman JJ (1982). *Case-control studies - design, conduct, analysis*, Oxford, UK: Oxford University Press.
- Swinscow TDV, revised by MJ Campbell (1998). *Statistics at square one* (11th ed.), London, UK: British Medical Association.

Notas del instructor

Módulo 28 SELECCIÓN DE UNA PRUEBA DE SIGNIFICACIÓN

Tiempo estimado y recursos didácticos

½ hora *	Introducción y discusión
1 hora*	Trabajo en equipo

* debe añadirse de ½ hora a 1 hora –a la sesión plenaria o al trabajo en equipo–, si el capacitador les pide a los participantes que completen los ejercicios sobre la selección de una prueba de significación en alguno de estos dos componentes de la sesión.

Introducción y discusión

- Este módulo no necesariamente debe presentarse en su totalidad. Como se indica en los objetivos, el principal propósito de la sesión es que los

participantes entiendan *qué* son las pruebas de significación y *por qué* deben emplearlas. Deben ser capaces de utilizar los diagramas de flujo para elegir las pruebas de significación adecuadas para los diferentes diseños de estudio y los distintos tipos de datos. *No* es necesario presentar todos los ejemplos que se dieron para el uso de los dos diagramas de flujo.

- La cuestión de *por qué* hacer pruebas de significación puede abordarse mediante la presentación de una tabla de dos por dos (por ejemplo, el número de fumadores y de no fumadores entre los hombres y las mujeres). Después se puede preguntar cómo puede interpretarse la diferencia entre hombres y mujeres (30% versus 20%). Sería mejor aún tomar una tabla de dos por dos de alguno de los propios equipos de investigación y hacer la misma pregunta.
- Una vez que a todos los participantes les haya quedado clara la justificación de las pruebas de significación, puede pedirles ejemplos de los resultados (tablas de dos por dos) de sus propios proyectos, para los que deban hacer pruebas de significación.
- Subraye que hay dos pruebas que son las más utilizadas: la prueba t y la prueba χ^2 . Las demás pruebas mencionadas en el diagrama de flujo (ilustración 28.1) tienen menos probabilidades de utilizarse en los tipos de proyectos más comunes.
- Si su grupo de participantes está lo suficientemente avanzado como para aprender por su cuenta a utilizar el diagrama de flujo y las matrices, pídale que completen algunos de los cinco ejercicios –o todos ellos; que se tomen unos minutos durante la sesión plenaria y empleen el diagrama de flujo y las matrices para seleccionar las pruebas adecuadas; pida enseguida voluntarios para responder. O bien, varios de los ejercicios –o todos ellos– pueden ser la primera tarea del trabajo en equipo. El facilitador tendría un papel activo para ayudar a los miembros del grupo a aficionarse al uso del diagrama de flujo y las tablas.
- Una vez que haya presentado este módulo, le *recomendamos que presente parte del siguiente: ya sea la prueba t o la prueba chi-cuadrada*. Esto hará más concretos los conceptos teóricos comentados en este módulo.

Trabajo en equipo

Permita que los participantes decidan para qué tabulaciones cruzadas deben hacer pruebas de significación, y determinen cuál es la prueba apropiada para cada una de las tabulaciones cruzadas que seleccionen.

Pruebas sugeridas para los estudios de investigación dados en los ejercicios

Ejercicio 1:

El estudio tiene dos *muestras*. Nos interesan las diferencias significativas en la presión arterial diastólica entre el grupo de estudio –que está tomando el nuevo fármaco–, y el grupo de comparación –que está tomando un placebo. Los datos son *numéricos* (tabla 28.1 o ilustración 28.1).

Las muestras no están pareadas; por lo tanto la *prueba t* es la prueba apropiada.

Ejercicio 2:

Supongamos que tenemos dos *muestras*. Se están estudiando las *diferencias* para determinar su significación; el resultado o la variable dependiente estudiada es el número de muertes perinatales, es decir, datos *nominales* (ilustración 28.1 o la tabla 28.1). Dichos estudios se hacen en muestras grandes y éstas *no están apareadas*. Por consiguiente, la prueba es *chi cuadrado*.

Si queremos encontrar la fuerza de la *asociación*, debemos utilizar la tabla 28.2. Como se trata de *datos nominales*, debemos calcular la razón de probabilidades o riesgo relativo. Aquí también puede usarse la tabla de 2 x 2 construida para la prueba χ^2 .

Ejercicio 3:

Hay dos muestras cuyas diferencias se están analizando para averiguar si tienen significación estadística. La variable dependiente es el peso, que es numérica. (Utilice la ilustración 28.1 o la tabla 28.1). Hubo solamente una muestra de 16 pacientes, pero las mediciones se hicieron dos veces en cada paciente. Esta es, pues, una *muestra coincidente* (o *apareada*). En este caso, se auto-pareó.

Por lo tanto la prueba es una *prueba t coincidente*. En cuanto a si la pérdida de peso se relaciona con el peso inicial, la prueba es del grado de *asociación*.

Puesto que los datos son *numéricos*, se calcula el *coeficiente de correlación de Pearson* (*r*). Si es necesario probar la significación estadística de la asociación, utilice la tabla 28.2, que lleva a la significación estadística de Pearson (*r*).

Ejercicio 4:

En este estudio se ha seleccionado una *muestra*. Se compara con datos de la población conocidos. La prueba debe abordar las diferencias entre las proporciones o porcentajes. Dado que los datos son *nominales*, se va a emplear la prueba χ^2 . (Véase la ilustración 28.1 o la tabla 28.1).

Este espacio en blanco intencionalmente dejado de la página

Módulo 29

**DETERMINACIÓN DE LAS DIFERENCIAS
ENTRE LOS GRUPOS:**

**PARTE I
ANÁLISIS DE LAS OBSERVACIONES SIN APAREAR**

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 29 DETERMINACIÓN DE LAS DIFERENCIAS ENTRE LOS GRUPOS

PARTE I: ANÁLISIS DE LAS OBSERVACIONES SIN APAREAR

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Decidir* cuándo aplicar la prueba t para observaciones sin aparear y la prueba de chi cuadrado.
2. *Calcular* los valores de t y los valores de chi cuadrado (χ^2).
3. *Usar* las tablas de t y las tablas de chi cuadrado (χ^2) para evaluar si el valor calculado de t y el valor calculado de chi cuadrado (χ^2) son significativos.
4. *Tomar una decisión* respecto de si puede usar estas pruebas en sus datos y, de ser así, qué prueba debe usar con qué datos.
5. *Hacer* estas pruebas con sus datos.

I. Introducción

II. Prueba t

III. Prueba de chi cuadrado (χ^2)

I. INTRODUCCIÓN

Al describir y analizar las tabulaciones cruzadas de sus principales variables (véanse los módulos 22 y 24), probablemente observó usted diferencias entre los grupos. Quizás ahora quiera saber si estas diferencias se deben a la casualidad o son reales (estadísticamente significativas).

Para determinarlo hay dos tipos de pruebas:

- la prueba *t*
- la prueba de chi-cuadrado (χ^2)

La *prueba t* se usa con datos NUMÉRICOS, cuando se compara la media de dos grupos.

La *prueba de chi cuadrado* se utiliza con datos CATEGÓRICOS, cuando se comparan las proporciones de acontecimientos que ocurren en dos o más grupos.

Ambas pruebas se usan con observaciones sin aparear. Con las observaciones APAREADAS se utilizan dos pruebas diferentes, dependiendo también de si los datos son categóricos o numéricos. (Véase el módulo 30).

II. PRUEBA T

La prueba *t* de Student se utiliza con datos numéricos para determinar si la diferencia entre las medias de dos grupos puede considerarse estadísticamente significativa.

Ejemplo 1:

En cierta provincia se ha observado una proporción muy alta de mujeres que paren por cesárea, así que se hace un estudio para averiguar a qué se debe. Puesto que se sabe que la baja estatura es uno de los factores de riesgo relacionados con los partos difíciles, el investigador se propone averiguar si en esta provincia hay una diferencia entre la altura media de las mujeres que tuvieron un parto normal y la de las que lo tuvieron por cesárea. La hipótesis nula sería que no hay diferencia en la altura media de los dos grupos de mujeres. Supongamos que se encuentran los siguientes resultados:

Tabla 29.1 Altura media de las mujeres con parto normal y las mujeres con parto por cesárea

<i>Tipo de parto</i>	<i>Número de mujeres incluidas en el estudio</i>	<i>Media de la estatura en cm</i>	<i>Desviación estándar</i>
Parto normal	60	156	3,1
Cesárea	62	154	2,8

La prueba t sería la apropiada para determinar si la diferencia de 2 cm observada puede considerarse estadísticamente significativa.

Para hacer la prueba t debe completar los siguientes tres pasos:

1. Calcular el valor de t
2. Elegir el nivel de significación y usar la tabla t
3. Interpretar los resultados

Paso 1. Cálculo del valor de t

Para calcular el valor de t debe usted seguir estos pasos:

- (1) *Calcule la diferencia entre las medias.* En el ejemplo anterior, la diferencia es: $156 - 154 = 2$ cm.
- (2) *Calcule la desviación estándar para cada grupo.* (El concepto de la desviación estándar y cómo se calcula se trató en el *módulo 27*). Suponga que se encontraron las desviaciones estándar que se muestran en la última columna de la *tabla 29.1*.
- (3) *Calcule el error estándar de la diferencia entre las dos medias.*

El error estándar de la diferencia se obtiene con la siguiente fórmula:

$$\sqrt{\frac{SD_1^2}{n_1} + \frac{SD_2^2}{n_2}}$$

Donde: SD_1 es la desviación estándar de la primera muestra
 SD_2 es la desviación estándar de la segunda muestra
 n_1 es el tamaño de la muestra de la primera muestra
 n_2 es el tamaño de la muestra de la segunda muestra

Con nuestros datos, si tomamos a las mujeres de parto normal como la muestra 1 y a las de cesárea como la muestra 2, el error estándar de la diferencia es:

$$\sqrt{\frac{3,1^2}{60} + \frac{2,8^2}{52}} = 0,56$$

- (4) Finalmente, *divida esa diferencia entre las dos medias, entre el error estándar de la diferencia.* El valor que se obtiene de esta manera se llama valor t.

En el ejemplo anterior: $t = \frac{2}{0,56} = 3,6$

Expresado en una sola fórmula:

$$t = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{SD_1^2}{n^1} + \frac{SD_2^2}{n^2}}}$$

(donde $\overline{X_1}$ es el valor de la media de la primera muestra, y $\overline{X_2}$ es el valor de la media de la segunda muestra)

Paso 2. Uso de una tabla t

Una vez que ha calculado el valor t, consulte la tabla t en el *anexo 29.1*, con la que podrá determinar si la hipótesis nula es rechazada o no.

- (1) En primer lugar, decida qué *nivel de significación* (valor- α o valor alfa) desea utilizar (vea el *módulo 28*). Recuerde que el nivel de significación elegido (valor- α) es una expresión de la probabilidad de encontrar una diferencia por casualidad, cuando no hay ninguna diferencia real. Normalmente elegimos un nivel de significación de 0,05.
- (2) En segundo lugar, determine el número de *grados de libertad* para hacer la prueba. Los grados de libertad son una medida derivada del tamaño de la muestra, que ha de tomarse en cuenta al realizar una prueba t. Cuanto mayor sea el tamaño de la muestra (y los grados de libertad), menor será la diferencia necesaria para rechazar la hipótesis nula.
- (3) En tercer lugar, ubique en la tabla el valor t correspondiente al valor- α (el nivel de significación que eligió) y los grados de libertad.

Si el valor de t que calculó es *igual* o *mayor* que el valor derivado de la tabla, entonces el valor p de significación es *menor* que el nivel- α que eligió (indicado en la parte superior de la columna). Luego, *rechazamos* la *hipótesis nula* y concluimos que hay una diferencia estadísticamente significativa entre las dos medias.

Si el valor de t que se calculó es *menor* que el valor derivado de la tabla, entonces el valor p es *mayor* que el nivel- α que eligió. Luego, *aceptamos* la *hipótesis nula* y concluimos que la diferencia observada no es estadísticamente significativa.

La manera de calcular el número de grados de libertad difiere de una prueba estadística a otra. Para la *prueba t de Student*, el número de grados de libertad se calcula como la suma de los dos tamaños de la muestra menos 2.

Así, al comparar en el *ejemplo 1* la estatura de las mujeres a las que se les hizo cesárea y las de parto normal, el número de grados de libertad es:

$$\text{g.l.} = 60 + 52 - 2 = 110$$

Nota:

Ésta es una manera aproximada de determinar los grados de libertad. Para ver el método exacto, consulte un libro de texto de estadística.

En nuestro ejemplo, buscamos el valor t correspondiente a $\alpha = 0,05$ y g.l. = 120, con lo que llegamos a 1,98.

Paso 3. Interpretación del resultado

Comparemos ahora el valor absoluto del valor t calculado en el paso 1 (es decir, el valor t , sin considerar el signo) con el valor t derivado de la tabla en el paso 2.

En nuestro ejemplo, el valor t calculado en el paso 1 es de 3.6, que es mayor que el valor t derivado de la tabla en el paso 2 (1.98). Así, el valor p es menor de 0.05. Luego, rechazamos la hipótesis nula y concluimos que la diferencia de 2 cm observada entre la altura media de las mujeres que tuvieron un parto normal y aquellas que lo tuvieron por cesárea es estadísticamente significativa.

Esta conclusión puede expresarse de diferentes maneras:

- Podemos decir que la probabilidad de que la diferencia de 2 cm de estatura observada entre los dos grupos de mujeres se deba a la casualidad es menor de 5%.
- También podemos decir que la diferencia entre los dos grupos es 3.6 veces el error estándar.

Si desea comparar el valor de la media de más de dos grupos (por ejemplo, la estatura de las mujeres de las zonas urbanas, semi-urbanas y rurales) no emplee la *prueba t de Student*. En este caso utilice la *prueba F*, que no se describe aquí.

III. PRUEBA DE CHI CUADRADO (χ^2)

Si usted tiene datos categóricos, la prueba chi-cuadrado se utiliza para averiguar si las diferencias observadas entre las proporciones de un acontecimiento dado en dos o más grupos pueden considerarse estadísticamente significativas.

Ejemplo 2:

Supongamos que en un estudio transversal de los factores que afectan el uso de las clínicas prenatales, encuentra que 64% de las mujeres que viven a no más de 10 kilómetros de la clínica acudieron para recibir atención prenatal, en comparación con sólo 47% de las que viven a más de 10 kilómetros de distancia. Esto sugiere que las mujeres que viven cerca de las clínicas utilizan los cuidados prenatales (CPN) con mayor frecuencia. Los resultados completos se presentan en la *tabla 29.2*.

Tabla 29.2 Uso de las clínicas prenatales por las mujeres que viven lejos y las que viven cerca de la clínica

<i>Distancia de los CPN</i>	<i>Uso de los CPN</i>	<i>No usaron CPN</i>	<i>TOTAL</i>
Menos de 10 km	51 (64%)	29 (36%)	80 (100%)
10 km o más	35 (47%)	40 (53%)	75 (100%)
Total	86	69	155

A partir de la tabla concluimos que parece haber una diferencia en el uso de la atención prenatal entre quienes viven cerca y quienes viven lejos de la clínica (64% versus 47%). Ahora queremos saber si esta diferencia es estadísticamente significativa.

Para responder a esta pregunta, podemos utilizar la prueba de chi cuadrado. Esta prueba consiste en medir la diferencia entre las frecuencias observadas y las frecuencias esperadas *si la hipótesis nula* (es decir, *la hipótesis de que no hay diferencia*) fuera cierta.

Para realizar una prueba χ^2 debe usted seguir los siguientes 3 pasos:

1. Calcular el valor de χ^2
2. Usar una tabla de χ^2
3. Interpretar χ^2

Paso 1. Cálculo del valor de χ^2

- (1) Calcule la frecuencia esperada (E) para cada celda.

Para encontrar la *frecuencia esperada E de una celda*, multiplique el total de la fila por el total de la columna y divídalo entre el gran total (general):

$$E = \frac{\text{total de la fila} \times \text{total de la columna}}{\text{gran total (general)}}$$

- (2) Para cada celda, reste la frecuencia esperada de la frecuencia observada ($O - E$).
- (3) Para cada celda, obtenga el cuadrado del resultado de ($O - E$) y divídalo entre la frecuencia esperada E .
- (4) Sume los resultados al cuadrado calculados en el paso (3) para todas las celdas.

La fórmula para calcular un valor de chi cuadrado (pasos (b) a (d)) es como sigue:

$$\chi^2 = \sum \frac{(O - E)^2}{E}$$

En donde: O Es la frecuencia observada (indicada en la tabla)

E es la frecuencia esperada (por calcular), y

Σ (la suma de) lo dirige para sumar los valores de $(O - E)^2 / E$ para todas las celdas de la tabla.

Para una tabla de dos por dos (que contiene 4 celdas) la fórmula es:

$$\chi^2 = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$$

Paso 2. Uso de una tabla χ^2

En cuanto a la prueba t, el valor calculado de χ^2 debe compararse con un valor teórico de χ^2 para determinar si la hipótesis nula es rechazada o no. El *anexo 29.2* contiene una tabla de valores teóricos de χ^2 .

- (1) En primer lugar, decida qué nivel de significación desea utilizar (el valor α o alfa). Normalmente tomamos 0,05.
- (2) Enseguida calcule los grados de libertad. Con la prueba χ^2 el número de grados de libertad se relaciona con el número de celdas, es decir, el número de grupos que está comparando. El número de grados de libertad se determina multiplicando el número de filas (r) menos 1 por el número de columnas (c) menos 1:

$$g.l. = (r-1) \times (c-1)$$

Para una tabla sencilla de dos por dos, el número de grados de libertad es 1:

$$g.l. = (2-1) \times (2-1) = 1$$

- (3) Después, debe ubicar en la tabla el valor de χ^2 correspondiente al valor α y el número de grados de libertad. Si el valor calculado de χ^2 es igual o mayor que el valor de χ^2 de la tabla, entonces el valor de p es menor que el valor (α) de significación que se eligió. En este caso, rechazamos la hipótesis nula y concluimos que hay una diferencia estadísticamente significativa entre los grupos. Si el valor calculado de χ^2 es menor que el valor de χ^2 de la tabla, entonces el valor de p es mayor que el nivel de significación que se eligió, de 0,05. En este caso, aceptamos la hipótesis nula y concluimos que la diferencia observada no es estadísticamente significativa.

Paso 3. Aplicación de χ^2

Al igual que en la prueba t, la hipótesis nula es rechazada si $p < 0,05$, lo que es el caso cuando el valor calculado de χ^2 es mayor que el valor teórico de χ^2 en la tabla.

Si aplicamos ahora la prueba de χ^2 a los datos del *Ejemplo 2* (uso de la atención prenatal), obtenemos el siguiente resultado:

Paso 1: Cálculo del valor de χ^2

En primer lugar, las *frecuencias esperadas* para cada celda se calculan de la siguiente manera:

$$E_1 = 86 \times 80/155 = 44,4$$

$$E_2 = 69 \times 80/155 = 35,6$$

$$E_3 = 86 \times 75/155 = 41,6$$

$$E_4 = 69 \times 75/155 = 33,4$$

Por razones de conveniencia las frecuencias observadas y esperadas se muestran en la siguiente tabla:

Tabla 29.3 Frecuencias observadas y frecuencias esperadas en el uso de las clínicas de atención prenatal

<i>Distancia de la APN</i>	<i>Uso de la APN</i>	<i>No uso la APN</i>	<i>Total</i>
Menos de 10 km	O1 = 51, E1 = 44.4	O2 = 29, E2 = 35.6	80
10 km o más	O3 = 35, E3 = 41.6	O4 = 40, E4 = 33.4	75
Total	86	69	155

Tome en cuenta que las frecuencias esperadas se refieren a los valores que habríamos esperado, dado el número total de 80 y 75 mujeres en los dos grupos, *si la hipótesis nula*, –según la cual *no* hay diferencia entre los dos grupos–, *fuera verdadera*.

Ahora puede calcularse *el valor de χ^2* :

$$\begin{aligned}\chi^2 &= \frac{(51 - 44,4)^2}{44,4} + \frac{(29 - 35,6)^2}{35,6} + \frac{(35 - 41,6)^2}{41,6} + \frac{(40 - 33,4)^2}{33,4} \\ &= 0,98 + 1,22 + 1,05 + 1,30 = 4,55\end{aligned}$$

Paso 2: Uso de la tabla de χ^2

Puesto que tenemos una sencilla tabla de dos por dos, el número de grados de libertad (g. l.) es 1.

Use la tabla de valores de chi cuadrado del *anexo 29,2*. Determinamos de antemano un nivel de significación de 5% (valor $\alpha = 0,05$).

Puesto que el número de g. l. es 1, buscamos sobre esa fila la columna en que $p = 0,05$. Esto nos da un valor de 3,84. Nuestro valor de 4,55 es *mayor* que 3,84, lo que significa que el valor de p es *menor* a 0,05.

Paso 3: Interpretación del resultado

Podemos entonces concluir que *las mujeres que viven a 10 km de la clínica utilizan la atención prenatal con una frecuencia significativamente mayor que las mujeres que viven a más de 10 km de distancia*.

Es importante que en el informe final de su estudio presente los datos con claridad y formule cuidadosamente cualesquiera conclusiones que se basen en las pruebas estadísticas.

Para el ejemplo anterior, podría presentar en el informe la *tabla 29.2* y plantear sus conclusiones de la siguiente manera

“La tabla 2 indica que 64% de las mujeres que viven a una distancia de 10 km de la clínica usaron atención prenatal durante el embarazo, en comparación con sólo 47% de las mujeres que viven a 10 km o más de la clínica más cercana. Esta diferencia es estadísticamente significativa ($\chi^2 = 4,55$; $p < 0,05$).”

Nota:

- La prueba de χ^2 sólo puede aplicarse si la muestra es bastante grande. La regla general es que el total de la muestra debe ser de al menos 40 y las *frecuencias esperadas* en cada una de las celdas deben ser de al menos 5. Si éste no fuera el caso, debería emplearse la prueba exacta de Fisher. (Véase el capítulo 9 del libro de Swinscow, *Statistics at Square One*, citado en este módulo.) Si la tabla tuviera más de dos filas por dos columnas, se aceptaría una frecuencia esperada de 1 de cada 5 celdas inferior a 5.

- A diferencia de la prueba t, la prueba de χ^2 también puede emplearse para comparar más de dos grupos. En este caso, se diseña una tabla de tres o más filas o columnas, en lugar de una tabla de dos por dos.

En el ejemplo anterior, se pueden distinguir tres diferentes distancias: menos de 5 km, de 5 a 10 km y más de 10 km. Los datos se pondrían entonces en una tabla de dos por tres. El número de grados de libertad sería $(3-1) \times (2-1) = 2$.

Formula rápida

Para las tablas de dos por dos, hay un método rápido para calcular el valor de chi cuadrado que puede reemplazar el paso 1 descrito anteriormente.

Si los diversos números de la tabla cruzada se representan con las siguientes letras:

	<i>Condición</i>		<i>Total</i>
	<i>+</i>	<i>-</i>	
Exposición			
Sí	a	b	E
No	c	d	F
Total	G	H	N

Donde, $E = a + b$; $F = c + d$; $G = a + c$; $H = b + d$,

La fórmula rápida para calcular el valor de chi cuadrado para una tabla de dos por dos es:

$$\chi^2 = \frac{N(ad - bc)^2}{(a + b)(c + d)(a + c)(b + d)} = \frac{N(ad - bc)^2}{EFGH}$$

Nota:

Las computadoras son útiles cuando se trata de grandes conjuntos de datos. Hay toda una variedad de *software* para hacer pruebas estadísticas, incluyendo los valores de *p*. Una *calculadora de estadística* también le da los valores de chi cuadrado.

Trabajo en equipo

Si recogió datos por observaciones sin aparear, identifique la prueba de significación apropiada y haga los análisis necesarios.

Anexo 29.1 Distribución de la t de Student

La primera columna muestra el número de grados de libertad. Los encabezados de las demás columnas dan los valores α para t para superar el valor de la entrada. Utilice la simetría en los valores negativos.

Grados de libertad	Valor de t si se elige $p(\alpha) = 0,05$	Valor de t si se elige $p(\alpha) = 0,01$
1	12,71	63,66
2	4,30	9,92
3	3,18	5,84
4	2,78	4,60
5	2,57	4,03
6	2,45	3,71
7	2,36	3,50
8	2,31	3,36
9	2,26	3,25
10	2,23	3,17
11	2,20	3,11
12	2,18	3,05
13	2,16	3,01
14	2,14	2,98
15	2,13	2,95
16	2,12	2,92
17	2,11	2,90
18	2,10	2,88
19	2,09	2,86
20	2,09	2,85
21	2,08	2,83
22	2,07	2,82
23	2,07	2,81
24	2,06	2,80
25	2,06	2,79
30	2,04	2,76
40	2,02	2,70
60	2,00	2,66
120	1,98	2,62
Infinito	1,96	2,58

Si el valor calculado de t (sin considerar el signo) es *mayor* que el valor indicado en la tabla, el valor de p en los cálculos de usted es *menor* que el valor elegido de p (α -), indicado en el encabezado de la columna.

En ese caso, la hipótesis nula, que establece que no hay ninguna diferencia, se *rechaza*, y se puede concluir que *sí hay* una diferencia significativa en el resultado de su estudio.

Anexo 29.2 Tabla de valores de χ^2

<i>Grados de libertad</i>	<i>Valor de χ^2 si $\alpha = 0.05$</i>	<i>Valor de χ^2 si $\alpha = 0.01$</i>
1	3.84	6.63
2	5.99	9.21
3	7.81	11.34
4	9.49	13.28
5	11.07	15.09
6	12.59	16.81
7	14.07	18.48
8	15.51	20.09
9	16.92	21.67
10	18.31	23.21
11	19.68	24.72
12	21.03	26.22

Si el valor calculado de χ^2 es *mayor* que el valor indicado en la tabla, el valor de p es *menor* que el nivel de significación seleccionado (indicado en el encabezado de la columna).

En ese caso, la hipótesis nula, que establece que no hay diferencia, es *rechazada*, y puede concluirse que la diferencia entre los dos grupos de estudio *es* estadísticamente significativa.

Anexo 29.3 Manejo de las variables de confusión: la prueba de chi cuadrado de Mantel-Haenszel

En la *tabla 29.4* se presentan los resultados de una investigación de la esquistosomiasis entre los habitantes de dos poblados.

Tabla 29.4 Prevalencia de la esquistosomiasis en dos poblados, A y B

	<i>Poblado A</i>		<i>Poblado B</i>		<i>Total</i>
Esquisto (+)	80	(32%)	80	(32%)	160
No esquisto (-)	170	(68%)	170	(68%)	340
Total	250	(100%)	250	(100%)	500

Al parecer, la prevalencia de la esquistosomiasis es la misma en ambos poblados (32%). No obstante, los investigadores sospecharon que la edad era una variable de confusión. Por consiguiente, la tabla 29.4 se dividió en dos (27.5 y 27.6). Tome en cuenta que si sumamos los números de las tablas 29.5 y 29.6 esto nos da la tabla 29.4.

Tabla 29.5 Prevalencia de la esquistosomiasis en niños de 5 a19 años en los poblados A y B

	Poblado A		Poblado B		Total
Esquisto (+)	37	(62%)	73	(38%)	110
No esquisto (-)	23	(38%)	117	(62%)	140
Total	60	(100%)	190	(100%)	250

$\chi^2 = 9,08$; 1 grado de libertad; $p < 0,01$.

Tabla 29.6 Prevalencia de la esquistosomiasis en personas de 20 años y mayores en los poblados A y B

	Poblado A		Poblado B		Total
Esquisto (+)	43	(23%)	7	(12%)	50
No esquisto (-)	147	(77%)	53	(88%)	200
Total	190	(100%)	60	(100%)	250

$\chi^2 = 2,78$; 1 grado de libertad; $p < 0,05$.

De las *tablas 29.5 y 29.6* queda claro que:

- En cada uno de los grupos de edad, la esquistisomiasis es más prevalente en el poblado A que en el B.
- La esquistosomiasis es más prevalente en los niños que en los adultos
- En el poblado A hay relativamente pocos niños y muchos adultos, en comparación con el poblado B.

Se dice que la edad es una variable de confusión porque está relacionada con la variable de interés (la prevalencia de la esquistosomiasis) y con los grupos comparados (residencia en el pueblo A o en el B).

Este ejemplo ilustra un punto importante del análisis de datos. Integrar datos distintos puede ser muy engañoso. En este ejemplo en particular, se estaba ocultando una importante diferencia real. En otras situaciones, integrar los datos puede sugerir una diferencia o una asociación que no existe, o incluso una diferencia opuesta a la que existe.

Por consiguiente, es importante analizar por separado los datos anteriores para los diferentes grupos de edad. Los valores adecuados de χ^2 (con corrección de continuidad) para

comparar la prevalencia en los poblados A y B se muestran en las *tablas 29.5 y 29.6*. La diferencia en la prevalencia es significativa para los niños pero no para los adultos.

La prueba de χ^2 de Mantel-Haenszel

Muchas veces es útil disponer de una prueba de síntesis que integre la evidencia de las tablas individuales, pero considere el factor de confusión (la edad, en nuestro ejemplo). Para hacer esto, se describe a continuación la prueba χ^2 de Mantel-Haenszel.

Para cada una de las tablas dos por dos vamos a usar la notación:

Exposición	Presente (+)	Ausente (-)	Total
Sí	a	b	E
No	c	d	F
Total	G	H	N

Paso 1. Para cada una de las tablas de dos por dos,

- 1. encuentre la frecuencia observada O_a
- 2. calcule la frecuencia esperada E_a , que es igual a EG/N
- 3. calcule la varianza V_a , que es igual a $EFGH/(N(N-1))$

Paso 2. El valor de chi cuadrado de Mantel-Haenszel (χ^2_{MH}) es

$$\chi^2_{MH} = \frac{(O - E - 0,5)^2}{V_a} \text{ con grados de libertad} = n1$$

Donde: O = la suma de las frecuencias observadas (O_a)
 E = la suma de las frecuencias esperadas (E_a)
 V = la suma de las varianzas (V_a)
0,5 es el factor de corrección de continuidad

Para comprobar la significación estadística, empleamos las tablas χ^2 , como se comentó anteriormente.

Aplicación:

En la prevalencia de la esquistosomiasis en los dos poblados, hay dos tablas de 2 por 2, una para los menores de 20 años (niños) y otra para las personas de 20 años y mayores (adultos). (Véanse las tablas 29.5 y 29.6). Las frecuencias observadas y las esperadas en las dos tablas se indican a continuación.

En el *ejemplo*, los cálculos son los siguientes:

	O_a	$E_a = EG/N$	$V_a = EFGN/(N^2(N-1))$
Niños	37	26,4	$110 \times 140 \times 60 \times 190/2502 \times 249 = 11,3$
Adultos	43	38	$50 \times 200 \times 190 \times 60/2502 \times 249 = 7,3$
Total	80	64,4	18,6

$O = 80, E = 64,4, V = 18,6$

$$\chi^2_{MH} = \frac{(80 - 64,4 - 0,5)^2}{18,6} = \frac{15,1^2}{18,6} = 12,25 \text{ (} p < 0,001 \text{)}$$

Puede concluirse, por consiguiente, que la prevalencia de la esquistosomiasis es significativamente diferente en los poblados A y el B. (Recuerde que éste parecía no ser el caso cuando revisamos la *tabla 29.4*, en que se integraron los datos de los adultos y los niños).

Validez de la prueba χ^2 de Mantel-Haenszel

La prueba χ^2 de Mantel-Haenszel es una prueba aproximada. La norma para evaluar su adecuación es más complicada que la prueba común de χ^2 . Para cada tabla, se calculan dos valores adicionales y se suman sobre las tablas. Éstos son:

- 1. $\min(E,G)$, que son la E y la G más pequeñas
- 2. $\max(O, G - F)$, que es O si G es menor o igual a F, y $F - G$ si G es mayor.

Estas dos sumas deberían diferir del total de los valores esperados, E_a , por al menos 5. Los pormenores del cálculo para el ejemplo anterior son:

	min (E,G)	max (O, G – F)
Niños	60	0
Adultos	50	0
Total	110	0

Estas sumas son 110 y 0, ambas difieren de 64.4 (E_a) por más de 5. Por lo tanto el uso de la prueba de Mantel-Haenszel es válido.

Módulo 29

**DETERMINACIÓN DE LAS DIFERENCIAS ENTRE
LOS GRUPOS,
PARTE I: ANÁLISIS DE LAS OBSERVACIONES
SIN APAREAR**

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
3 horas +	Trabajo en equipo

Introducción y discusión

- Le recomendamos que *presente ya sea la prueba t o la prueba de χ^2 inmediatamente después de completar el módulo 28, de manera que los participantes puedan tener una mejor idea de lo que es una prueba de significación y cómo se usa*. Quizás todos los equipos de investigación apliquen pruebas χ^2 a sus datos, pero no todos van a utilizar la prueba t. Por lo tanto, de preferencia presente la prueba de χ^2 en combinación con el módulo 28 y la prueba t en otra sesión.
- Proceda despacio, paso a paso, al explicar cómo funcionan las pruebas de significación, para no asustar a los participantes de poca experiencia. Subraye que no es importante entender por qué los cálculos de los valores de t y de χ^2 se hacen de esta manera (en realidad, no siempre hay una justificación clara; tómese como ejemplo el concepto de grados de libertad). Basta saber cómo se hacen. Tenga cuidado con las fórmulas: preséntelas después de explicar paso a paso los cálculos.
- Si lo prefiere, use como ejemplo los propios estudios del grupo en lugar de los ejemplos 1 y 2 presentados en el módulo. Recuerde, sin embargo, utilizar ejemplos sencillos, es decir, de tablas de dos por dos y números pequeños, de modo que sean más fáciles de seguir.
- Tome precauciones adicionales para explicar cómo utilizar la tabla de t y la tabla de χ^2 y cómo interpretar los resultados. Permita que los participantes luchen con los ejemplos antes de darles las respuestas correctas.
- También preste atención a que la formulación de las conclusiones basadas en las pruebas de significación sea adecuada, tanto con los resultados significativos como con los que no lo son.

- *El anexo 29.3* no debe presentarse a menos que el grupo esté muy adelantado. De ser necesario, los equipos pueden leerlo y usar la prueba.

Trabajo en equipo

Cuando haga pruebas estadísticas cerciórese de que cada uno de los miembros del grupo haga por lo menos una prueba por sí solo.

Este espacio en blanco intencionalmente dejado de la página

Módulo 30

DETERMINACIÓN DE LAS DIFERENCIAS ENTRE LOS GRUPOS:

PARTE II ANÁLISIS DE LAS OBSERVACIONES APAREADAS¹

¹ La mayor parte de este módulo se deriva del material del curso de Maestría de la London School of Hygiene and Tropical Medicine, con el permiso de Richard Hayes, Betty Kirkwood y Tom Marshall.

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 30 DETERMINACIÓN DE LAS DIFERENCIAS ENTRE LOS GRUPOS, PARTE II: ANÁLISIS DE LAS OBSERVACIONES APAREADAS

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Identificar* los estudios de investigación en los que sea necesario aparear o emparejar a los individuos.
2. *Identificar y usar* las pruebas de significación adecuadas para los estudios que emplean datos apareados.

- I. Introducción
- II. Prueba t para datos apareados
- III. Prueba de chi cuadrado de McNemar

I. INTRODUCCIÓN

Este módulo describe las pruebas más comúnmente empleadas para las observaciones apareadas correspondientes a los datos numéricos y a los datos nominales:

- la *prueba t* para los datos numéricos apareados
- la *prueba de chi cuadrado de McNemar* para los datos nominales

Recordará usted del *módulo 9* y la parte III del *módulo 26*, que las observaciones apareadas o emparejadas se emplean cuando los investigadores quieren cerciorarse –mediante el diseño de su estudio– de que la relación entre dos variables –en las que están interesados– no se confunda con otra variable. Deben, por consiguiente, formar la muestra de sus casos y sus controles de tal manera que éstos sean semejantes respecto de una o más posibles variables de confusión.

El concepto de aparear o emparejar a los individuos se ilustra en los siguientes ejemplos:

Ejemplo 1:

Un investigador quiere averiguar si una clase de estudiantes a quienes se enseñó con asistencia audiovisual (AA) tiene, en promedio, mejores calificaciones que aquellos a quienes se enseñó sin dicho recurso. Para minimizar el efecto de variables de confusión tales como la *situación social* y el *conocimiento previo* de los estudiantes, cada estudiante de la clase con AA se pareó con otro de la clase sin AA de similar condición social y nivel de conocimientos.

Ejemplo 2:

En una investigación nutricional, se hizo un ejercicio de control de calidad para comprobar la concordancia entre dos observadores en la medición del peso de los niños. En este caso pareamos las observaciones ya que tenemos *un conjunto de dos observaciones del mismo niño*.

Ejemplo 3:

Un equipo de estudio comparó el conteo de huevos de esquistosomiasis en dos poblados. Reconoció que los conteos de huevos varían con la edad y el sexo. Para cerciorarse de que las muestras fueran comparables con respecto a estos dos factores, decidió seleccionar a los individuos por pares, cada par con un miembro de cada uno de los dos poblados, emparejados por *edad y sexo*.

II. PRUEBA T PARA DATOS APAREADOS

En el *módulo 29* se hizo una comparación de la media de las muestras para observaciones numéricas –sin aparear– de la estatura de las madres embarazadas, utilizando la prueba t. Cuando se trata de observaciones apareadas (emparejadas), la comparación de la media de las muestras se hace empleando una prueba t modificada, conocida como la *prueba t para datos apareados*.

En la *prueba t para datos apareados* las diferencias entre las observaciones apareadas se utilizan en lugar de los dos conjuntos originales de observaciones.

La *prueba t para datos apareados* calcula el valor de t como:

t = media de las diferencias de error estándar

$$t = \frac{\text{media de las diferencias}}{\text{error estándar}}$$

Los grados de libertad se obtienen con el número de observaciones apareadas (= tamaño de la muestra) menos 1.

Para interpretar el resultado del estudio, se usa la misma tabla de valores de t que se empleó en la prueba t para observaciones sin aparear (véase el anexo 29.1).

Para ilustrar cómo se utiliza la *prueba t para datos apareados*, empleamos los resultados de la encuesta de nutrición que acabamos de ver en el ejemplo 2. Los resultados (Tabla 30.1).

La hipótesis nula en este estudio es que si los observadores A y B midieran a todos los niños de la población de la que se tomó la muestra de estos 20 niños, no habría, en promedio, ninguna diferencia entre sus mediciones. En otras palabras, la media de la diferencia entre A y B sería cero.

Podemos considerar este conjunto de 20 diferencias (la columna A - B), como una muestra de todas las diferencias que se habrían obtenido si los observadores hubieran medido a toda la población.

Para hacer la prueba de significación, debe calcularse el valor de t y compararlo con el valor de t en la tabla, para determinar si hay entre los dos valores una diferencia estadísticamente significativa. Esto indica la probabilidad de que los resultados del estudio se hubieran dado por casualidad.

Tabla 30.1 Resultados del ejercicio de control de calidad durante una investigación nutricional

<i>Niño No.</i>	<i>Medidas de peso (kg)</i>		<i>Diferencia A – B (kg)</i>
	<i>Observador A</i>	<i>Observador B</i>	
1	18,6	17,7	0,9
2	17,1	14,5	2,6
3	14,3	12,4	1,9
4	23,2	20,7	2,5
5	18,4	16,8	1,6
6	14,9	14,4	0,5
7	16,6	14,1	2,5
8	14,8	17,1	-2,3
9	21,5	21,2	0,3
10	24,6	21,9	2,7
11	17,4	16,6	0,8
12	15,7	13,6	2,1
13	16,1	14,5	1,6
14	12,9	11,2	1,7
15	12,3	16,0	-3,7
16	19,4	20,4	-1,0
17	19,3	17,5	1,8
18	24,8	22,2	2,6
19	14,3	15,1	-0,8
20	13,4	10,9	2,5

La prueba de significación se realiza de la siguiente manera:

1. Calcule la media de la diferencia de las mediciones entre A y B en la muestra. Esto es la suma de las diferencias dividido entre el número de mediciones:

$$\text{Media de la diferencia} = \frac{21,1}{20} = 1,05$$

2. Calcule la desviación estándar de las diferencias (*módulo 27*):

$$\text{Desviación estándar} = 1,77$$

Calcule el error estándar (*módulo 27*):

$$\text{Error estándar} = \sqrt{\frac{\text{Desviación estándar}}{\text{Tamaño de la muestra}}} = \sqrt{\frac{177}{20}} = 0,04$$

3. El valor de *t* es la media de la diferencia dividida entre el error estándar:

$$t = \frac{1,05}{0,40} = 2,62$$

4. Vaya a la tabla de valores de *t* del *anexo 29.1*.

Los grados de libertad se obtienen del tamaño de la muestra (el número de pares de observaciones) menos 1, lo que en este caso es 20 – 1 = 19.

La probabilidad –a partir de la tabla– es < 0.05, lo que nos permite concluir que hay una diferencia significativa entre los observadores. A y B necesitan definitivamente una mayor capacitación y supervisión en sus habilidades para la medición puesto que la prueba no muestra si sólo uno de ellos o los dos fueron inexactos en sus mediciones.

III. LA PRUEBA DE CHI CUADRADO DE MCNEMAR

La *prueba de chi cuadrado de McNemar* se usa con datos NOMINALES para comparar PROPORCIONES de observaciones apareadas. Es importante señalar que la disposición de la tabla es diferente de la que se emplea con muestras sin aparear.

La *tabla 30.2* muestra los resultados de un estudio de casos y controles que se hizo para determinar las causas de un brote de cólera en Bombay. Por cada caso confirmado de cólera en el hospital, se buscó un individuo del mismo sexo, la misma década de edad y el mismo barrio.

Tabla 30.2 Fuente de agua potable por pares de pacientes de cólera y controles dentro de los 5 días anteriores a la enfermedad (disposición incorrecta)

	Casos	Controles	Total
Pozo superficial	42 (55%)	15 (20%)	57
Agua de grifo	34 (45%)	61 (80%)	95
Total	76 (100%)	76 (100%)	152

Sin embargo, la disposición de la *tabla 30.2* no es correcta, no toma en cuenta que los casos y los controles se seleccionaron como pares.

La disposición correcta se presenta en la *tabla 30.3*.

Tabla 30.3 Fuente de agua potable por pares de pacientes de cólera y controles dentro de los 5 días anteriores a la enfermedad (disposición correcta)

<i>Controles</i>	<i>Casos</i>				<i>Total</i>
	<i>Pozo superficial</i>		<i>Agua de grifo</i>		
Pozo superficial	12	(16%)	3	(4%)	15
Agua de grifo	30	(39%)	31	(41%)	61
Total	42		34		70 (100%)

Adaptada de Baine & Mazotti *et al.* (1973).

¿Cómo debemos interpretar la *tabla 30.3*?

Tenemos que 12 pares (tanto el caso como el control) sacaron agua de pozo superficial y 31 sacaron agua de grifo. Por lo tanto, estos 43 pares no nos dan información acerca de si sacar agua de pozos superficiales es un factor de riesgo para contraer el cólera o no. Sin embargo, en 30 pares (39%) los casos sacaron agua de pozos superficiales mientras que los controles sacaron agua de grifo, y en sólo 3 pares (4%) los controles sacaron agua de pozos superficiales, mientras que los casos sacaron agua del grifo. Parecería, entonces, que sacar agua de pozos superficiales es un factor de riesgo para contraer el cólera.

Antes de aceptar esta conclusión, debemos hacer una prueba de significación para estimar la probabilidad de que estos resultados se deban a la casualidad o sólo a la variación de muestra. La prueba de significación adecuada es la prueba de chi-cuadrado de McNemar (véase la tabla 28.1):

$$\chi^2 = \frac{(|r - s| - 1)^2}{r + s} \text{ con 1 grado de libertad}$$

Donde: r = El número de pares en que el control sacó agua de grifo y el caso de un pozo superficial

s = el número de pares en que el control sacó agua de un pozo superficial y el caso la sacó del grifo

$|r - s|$ significa la diferencia entre r y s como un número positivo, independientemente de si s es mayor que r

Para comprobar la significación estadística, usamos tablas de chi cuadrado comunes y corrientes (*anexo 29.2*).

Nota:

La prueba χ^2 de McNemar es válida sólo si $(r + s)$ es mayor que 10.

La prueba puede hacerse con los datos de nuestro ejemplo porque $r + s$ ($30 + 3$) es mayor que 10.

El cálculo del valor de chi cuadrado es como se muestra a continuación:

$$\chi^2 = \frac{(30 - 3 - 1)^2}{30 + 3} = \frac{26^2}{33} = 20,5 \text{ con 1 grado de libertad}$$

Usando un nivel α de 0.01, el valor de la tabla de χ^2 es igual a 6.63 (*anexo 29.2*). Vemos que el valor calculado de χ^2 de 20.5 es mayor que el valor de la tabla. Esto significa que el valor p es menor a 0.01. Luego, rechazamos la hipótesis nula y concluimos que sacar agua de pozos superficiales es un factor de riesgo para contraer el cólera.

TRABAJO EN EQUIPO

Si sus datos se recogieron por pares u observaciones apareadas, identifique la prueba estadística adecuada y haga los cálculos y el análisis necesarios.

Referencias

Revise los libros de texto de epidemiología u estadística recomendados en los *módulos 9 y 28*.

Baine WB, Mazotti M, Greco D *et al.* (1974). "Epidemiology of cholera in Italy in 1973". *Lancet* ii (Dec.): 1370–1374.

**Módulo 30 DETERMINACIÓN DE LAS DIFERENCIAS
ENTRE LOS GRUPOS,
Parte II: ANÁLISIS DE LAS OBSERVACIONES
APAREADAS**

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
2 horas	Trabajo en equipo

Introducción y discusión

- Si ninguno de los equipos tiene observaciones apareadas y los participantes tienen poca experiencia en estadística, no hace falta presentar este módulo.
- Al explicar la forma de calcular los valores de t y los valores de χ^2 , vaya paso a paso, muy despacio. De nuevo, es más importante que los participantes entiendan *cómo* hacer los cálculos, que por qué se hacen así.
- Tómese el tiempo necesario para que todos los participantes sepan cómo leer la *tabla 30.3*. Ésta puede ser la primera tabla con números que representan *pares* de observaciones a la que se enfrentan (si se saltaron la parte respectiva del *módulo 26*).
- Cerciórese de que los participantes sepan cómo usar la tabla de t -y la tabla χ^2 , y cómo interpretar los resultados.

Módulo 31

MEDICIÓN DE LAS ASOCIACIONES ENTRE LAS VARIABLES: REGRESIÓN Y CORRELACIÓN*

(Opcional)

* La mayor parte de este módulo proviene del material del curso MS de la London School of Hygiene and Tropical Medicine, con la autorización de Richard Hayes, Betty Kirkwood y Tom Marshall.

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 31 MEDICIÓN DE LAS ASOCIACIONES ENTRE LAS VARIABLES; REGRESIÓN Y CORRELACIÓN

OBJETIVOS

Al terminar esta sesión usted debe ser capaz de:

1. *Ilustrar* la relación entre dos variables numéricas en un diagrama de dispersión.
2. *Interpretar* una línea de regresión.
3. *Calcular* e interpretar un coeficiente de correlación.
4. *Hacer* una prueba de significación del coeficiente de correlación.

I. Introducción.

II. Los diagramas de dispersión.

III. Determinación de la relación lineal: ajuste de una línea de regresión.

IV. Coeficientes de correlación.

V. Prueba de significación de un coeficiente de correlación.

I. INTRODUCCIÓN

Al explorar las asociaciones entre las variables, debemos distinguir los datos nominales, los ordinales y los numéricos (véanse la ilustración 28.1 y la tabla 28.1 en el módulo 28).

El *módulo 25* se ocupó de las asociaciones entre los datos nominales, en particular en los estudios de casos y controles.

Para las asociaciones entre los datos ordinales –en cuyo caso la significación puede determinarse y probarse calculando el *coeficiente de correlación de Spearman* o el *tau de Kendall*–, puede consultar un libro de texto de estadística (véanse las referencias indicadas en el *módulo 28*).

En este módulo examinamos las asociaciones entre los datos numéricos de las que se sospeche que tienen una relación lineal.

II. DIAGRAMA DE DISPERSIÓN

El primer paso para analizar la relación entre dos variables numéricas, medidas en los mismos individuos, siempre es trazar un *diagrama de dispersión*.

Ejemplo 1:

En un estudio de nutrición en un gran distrito rural, se pesó una muestra de 20 niños de 5 años de edad, y se estimó el ingreso de sus familias. Los resultados fueron los siguientes:

Tabla 31.1 Peso e ingreso familiar de 20 niños de 5 años de edad

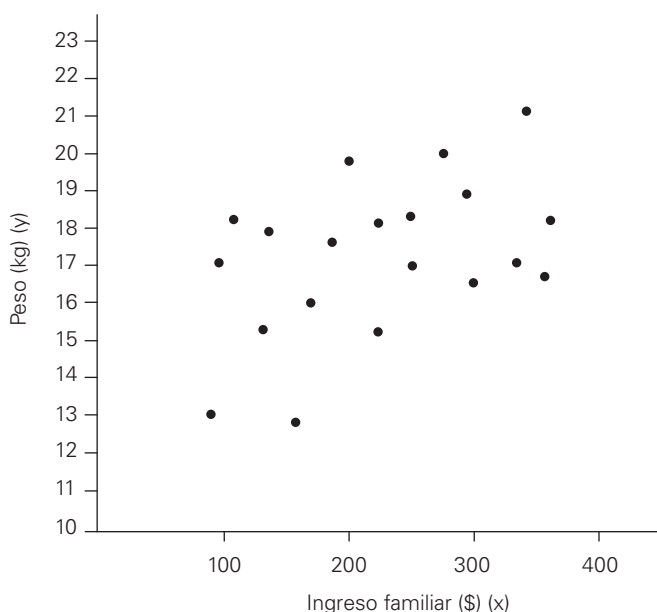
<i>Ingreso familiar en \$/año</i>	<i>Peso en kg</i>	<i>Ingreso familiar en \$/año</i>	<i>Peso en kg</i>
130	15,5	225	18,1
200	19,8	95	17,4
345	21,5	130	17,9
245	16,8	330	17,0
155	12,6	295	18,7
300	16,6	170	16,0
360	18,1	250	18,2
105	18,7	355	16,4
80	13,1	220	15,4
275	20,1	175	17,6

El objetivo era examinar si, para esta muestra de niños, el peso y el ingreso familiar estaban relacionados. Se puede dividir a los niños en dos categorías de ingresos, una categoría de ingreso elevado (por ejemplo, \$200 dólares o más), y una categoría de bajo ingreso (por ejemplo, menos de \$200 dólares), y calcular y comparar la media del peso en cada categoría para ver si hay diferencia. Para determinar si la diferencia es significativa debe usarse una prueba t. (Véase el módulo 29).

Hecho este análisis, se puede concluir que los niños de las familias de bajo ingreso tienen, en promedio, un peso más bajo que los niños de las familias de ingreso elevado. Sin embargo, sería más informativo tomar en cuenta *todas* las medidas individuales e investigar si las dos variables “ingreso de la familia” y “peso de los niños de cinco años” están asociadas.

Para ello se puede trazar el siguiente diagrama de dispersión:

Ilustración 31.1 Peso e ingreso familiar de 20 niños menores de 5 años



Notas sobre el trazo de un diagrama de dispersión:

1. Si estamos analizando cómo se asocia una variable dependiente con una variable independiente, por lo general ponemos la variable dependiente en el eje vertical (el eje “y”), y la variable independiente en el eje horizontal (el eje “x”). A veces no está claro cuál es la variable dependiente, en cuyo caso la elección del eje es arbitraria.

2. Seleccione las escalas de modo que la dispersión llene una parte razonable del diagrama.
3. Si un eje no empieza desde cero, muéstrelo claramente “rompiendo” el eje (como se hizo para el eje del peso en el ejemplo anterior).
4. Etiquete los ejes claramente.
5. Los puntos trazados deben ser lo suficientemente grandes como para que destaquen, de modo que sea fácil ver la dispersión.

III. DETERMINACIÓN DE LA RELACIÓN LINEAL: AJUSTE DE UNA LÍNEA DE REGRESIÓN

Si estamos interesados en determinar el peso del niño y conocemos el ingreso familiar, entonces el *peso* es la *variable dependiente* (y) y el *ingreso familiar* es la *variable independiente* (x). En el diagrama de dispersión anterior parece haber una tendencia al alza en el peso con el creciente ingreso familiar. Podemos trazar una línea a través de la dispersión de los puntos, como un sencillo resumen de la relación entre estas dos variables. Esto puede hacerse “a ojo” —una regla transparente sirve para hacerlo. Sin embargo, ajustar “a ojo” es bastante subjetivo y preferimos emplear una técnica más objetiva. El *método de mínimos cuadrados* da la “mejor” línea recta, utilizando la técnica descrita a continuación.

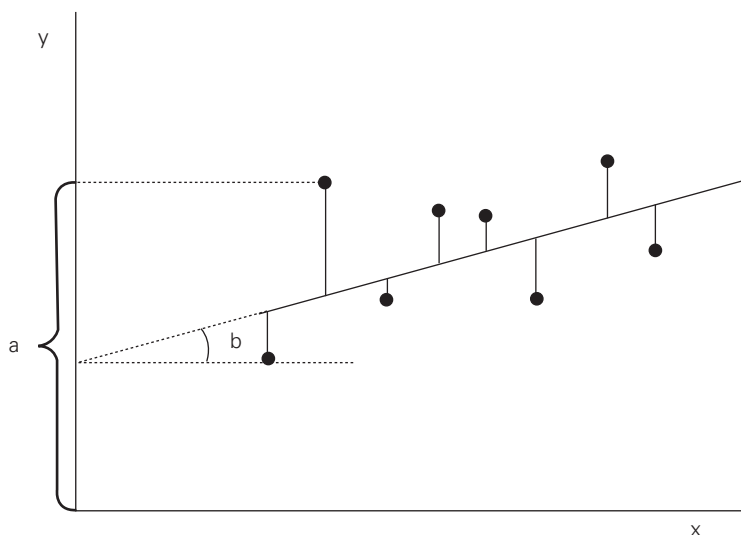
Cualquier línea recta trazada sobre un gráfico puede representarse con la ecuación:

$$y = a + bx$$

Cada punto de la línea tiene un valor “ x ” y un valor “ y ”; y la ecuación nos dice cómo están relacionados estos valores “ x ” y “ y ”. Los distintos valores de “ a ” y “ b ” dan diferentes líneas rectas. El valor de “ a ” nos da la INTERCEPTACIÓN de la línea en el eje “ y ” (“ a ” es la distancia desde cero hasta el punto donde la línea cruza el eje “ y ”; es decir, es el valor de “ y ” cuando el valor de “ x ” es cero) y “ b ” indica la PENDIENTE de la línea.

Para determinar la línea que se ajuste a nuestra dispersión, debemos decidir qué valores de “ a ” y “ b ” usar. Básicamente los escogemos de tal manera que las distancias verticales entre los puntos y la línea se minimicen. (Para ser más precisos, elegimos la “ a ” y la “ b ” que minimicen la suma de los cuadrados de estas distancias verticales. De ahí el nombre de “método de mínimos cuadrados”).

El *anexo 31.1* explica cómo calcular manualmente los valores de “ a ” y “ b ” a partir de los datos. Algunas calculadoras y programas de computadoras dan de manera automática los valores de “ a ” y “ b ”.



En nuestro ejemplo, utilizando una calculadora apropiada, encontramos:

$$a = 15,09 \quad b = 0,00984$$

De modo que la ecuación de nuestra línea ajustada es

$$y = 15,09 + 0,00984 \text{ veces } x$$

Para trazar esto en un diagrama de dispersión escogemos dos valores de "x", encontramos los valores de "y" correspondientes utilizando la ecuación, trazamos los dos puntos en el gráfico "y" y los unimos con una línea recta.

La línea ajustada se llama *regresión lineal* del peso y el ingreso familiar.

Por ejemplo:

$$x = 0 \quad \text{nos da} \quad y = 15.09$$

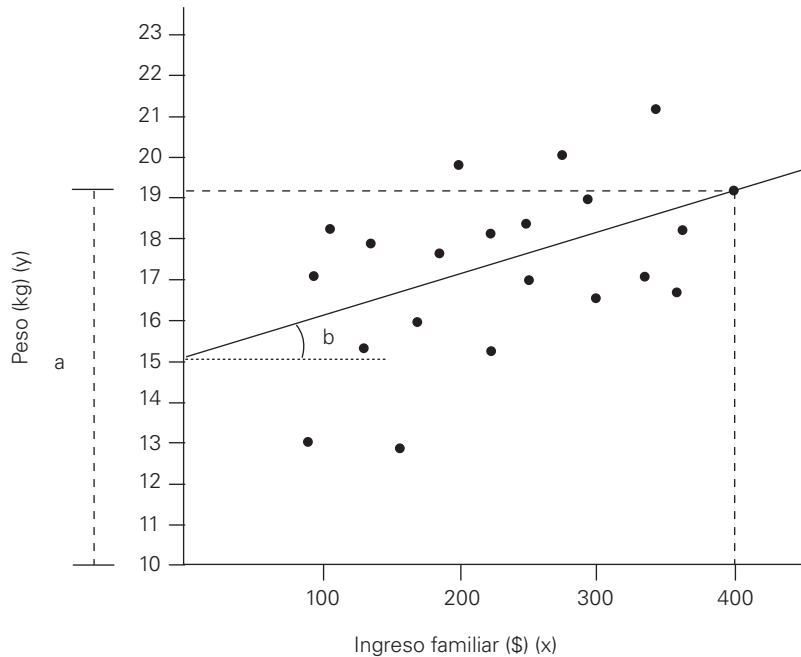
$$x = 400 \quad \text{nos da} \quad y = 15.09 + (0.00984 \times 400) = 19.03$$

Interpretación de una línea de regresión

La línea de regresión estima el valor medio de "y" para un valor dado de "x". Por ejemplo, nos dice que los niños cuyas familias tienen un ingreso de \$ 200 al año pesan alrededor de 17 kg en promedio, aunque algunos pesan más y algunos menos.

La pendiente b, llamada coeficiente de regresión, nos dice el incremento medio de "y" correspondiente a una unidad de crecimiento de "x". Así, en nuestro ejem-

Ilustración 31.2 Regresión lineal del peso de niños de 5 años y el ingreso familiar



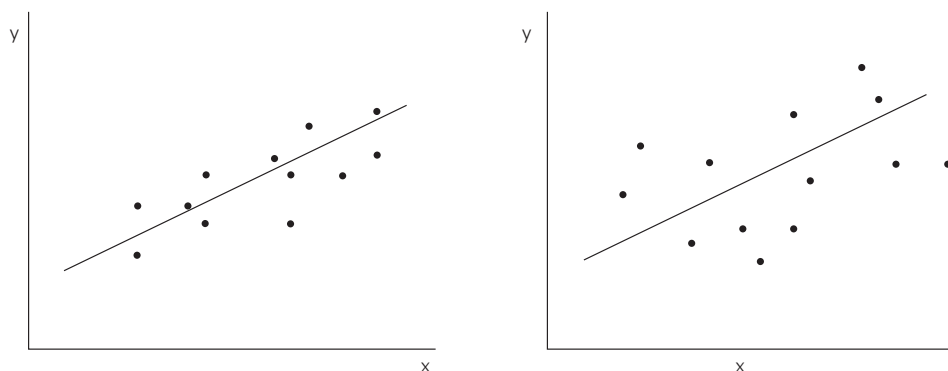
plo, el incremento del peso medio de 0,00984 kg (o alrededor de 10g) por cada incremento de 1 dólar en del ingreso familiar (o alrededor de 1 kg de peso por cada incremento de \$ 100).

Una nota de advertencia:

- Debe ajustarse una línea recta sólo si el diagrama de dispersión sugiere que la relación entre las dos variables es más o menos lineal. Se dispone de métodos más complejos para ajustar las curvas a los datos.
- Es peligroso extrapolar la línea de regresión fuera del recorrido (intervalo de valores) de los datos. Si en nuestro ejemplo extrapolamos la línea a un ingreso de \$2000 al año, ello arroja un peso medio estimado de 34,8 kg, lo que es absurdo.
- En la regresión, es importante tener claro cuál es la variable dependiente y cuál la independiente, ya que si se intercambian se obtiene una ecuación de regresión diferente.

IV. COEFICIENTES CORRELACIÓN

Considere los siguientes dos diagramas de dispersión:



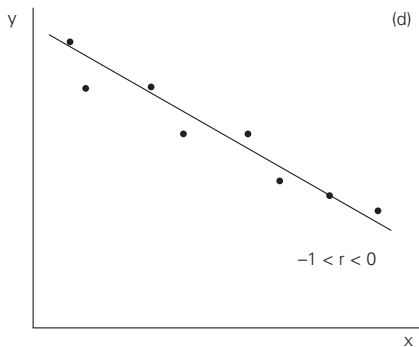
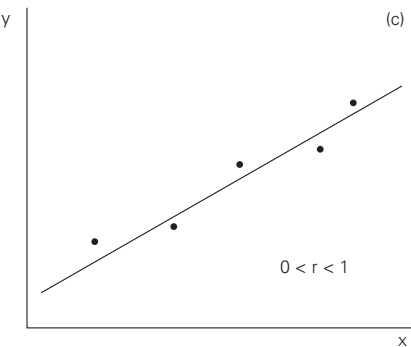
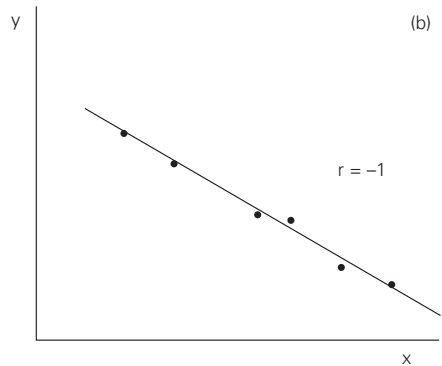
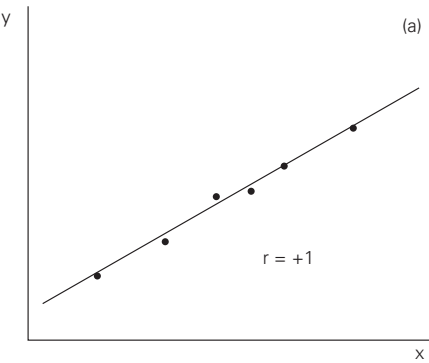
Los coeficientes de regresión b (es decir, las pendientes de las líneas) son idénticos en estos dos ejemplos, pero la dispersión en torno a la línea es mucho mayor en el segundo. Queda claro que la relación entre las variables “ y ” y “ x ” es mucho más estrecha en el primer diagrama.

Si sólo nos interesa *medir* la asociación entre las dos variables, entonces el *coeficiente de correlación de Pearson* (r) nos da una estimación de la fuerza de la asociación lineal entre dos variables numéricas.

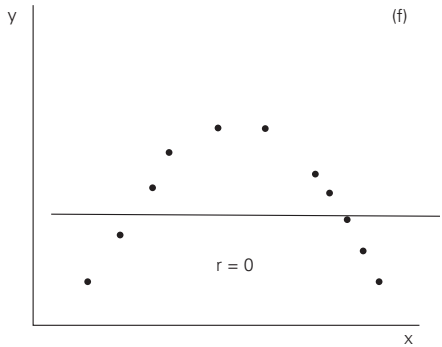
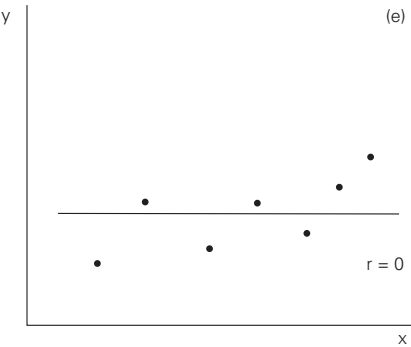
El coeficiente de correlación de Pearson puede calcularse manualmente (véase el anexo 31.2), o bien el valor de r puede obtenerse utilizando una calculadora programada o diversos programas de computación.

El coeficiente de correlación tiene las siguientes propiedades:

1. Para cualquier conjunto de datos, r se ubica entre -1 y $+1$.
2. Si $r = +1$, o -1 , la relación lineal es perfecta; esto es, todos los puntos se encuentran exactamente sobre una línea recta. Si $r = +1$, la variable “ y ” aumenta conforme aumenta “ x ” (es decir, sube la pendiente de la línea). (Véase el diagrama a.) Si $r = -1$, la variable “ y ” decrece conforme “ x ” aumenta (es decir, baja la pendiente de la línea). (véase el diagrama b).
3. Si r está entre 0 y $+1$, la línea de regresión tiende a subir, pero los puntos se dispersan alrededor de la línea. Cuanto más se acerca “ r ” a 1 , más se aproximan los puntos a la línea. (Véase el diagrama c). Lo mismo es cierto para los valores negativos de “ r ”; entre 0 y -1 , pero en este caso la pendiente de regresión tiende a bajarse. (Véase el diagrama d).



4. Si $r = 0$, no hay una relación *lineal* entre “ y ” y “ x ”. Esto puede significar que no hay relación alguna entre las dos variables (es decir, conocer el valor de “ x ” no nos dice nada del valor de “ y ”). (Véase el diagrama e). Sin embargo, también podríamos obtener $r = 0$ si hubiera una relación *curva* entre “ y ” y “ x ”. (Véase el diagrama f).



5. Una interpretación útil de r es que su cuadrado (r^2) mide la proporción de la *variabilidad* de la variable b , correspondiente a la relación lineal con la variable “ x ”.

Volviendo a nuestro ejemplo del peso y el ingreso familiar, la calculadora arroja:

$$r = 0,414$$

que es positiva (lo que indica una línea ascendente), pero lejos de 1 (lo que indica que hay mucha dispersión en torno a la línea).

V. PRUEBA DE LA SIGNIFICACIÓN DEL COEFICIENTE DE CORRELACIÓN

El valor de r se calculó a partir de una muestra de sólo 20 niños. Por consiguiente, el resultado está sujeto a error de muestreo y es poco probable que sea igual al verdadero valor de r , el cual obtendríamos si midiéramos a todos los niños de 5 años de este distrito.

Surge la pregunta de si realmente hay alguna relación entre el peso y el ingreso. Tal vez con la población total de niños de 5 años de edad, el diagrama de dispersión se vería como el *diagrama e* anterior (no hay relación entre "y" y "x") y la relación positiva en nuestra muestra se produjo por casualidad.

Para evaluar si éste es el caso, hacemos una prueba de significación con r . La hipótesis nula es que en la totalidad de la población no hay una relación lineal entre "y" y "x". Para hacer la prueba se calcula

$$t = r \times \sqrt{\frac{n-2}{1-r^2}}$$

Comparamos este valor de t con las tablas de distribución de t , con $(n - 2)$ grados de libertad, donde n es el número de observaciones.

En nuestro ejemplo: $n = 20$ $r = 0.414$

Por lo tanto $t = 0,414 \times \sqrt{\frac{18}{1-0414^2}} = 1,93$

Usando un valor de α (valor elegido de p) de 0,05, el valor de la tabla de t para 18 grados de libertad ($t_{18;0,05}$) = 2.10. (Véase el anexo 29.1). Así, el valor calculado de t es menor que el valor de la tabla; esto significa que el valor de p es mayor de 0.05; luego, la relación lineal *no* es estadísticamente significativa. Dado que el valor calculado es próximo al valor de la tabla (1.93; 2.10), éste es, de hecho, un caso "fronterizo".

Asociación y causalidad

Note que la existencia de una asociación estadística, incluso siendo muy fuerte, NO establece que un incremento en "x" *cause* un aumento en "y"; o que un incremento "y" cause un aumento en x. Una debilidad fundamental de los

estudios observacionales es que pueden demostrar la asociación, pero no la causalidad. Para demostrar una relación causal se tendría que elegir un diseño de estudio experimental.

Anexo 31.1 Ajuste manual de una línea de regresión en un diagrama de dispersión (el ingreso familiar y peso de niños de 5 años de edad)

Los valores de “a” (15.9) y “b” (0.00984) obtenidos directamente de la calculadora también pueden calcularse manualmente de la siguiente manera:

Para la ecuación de regresión

$$y = a + bx$$

$$b = \frac{\sum xy - (\sum x)(\sum y) / n}{\sum x^2 - (\sum x)^2 / n}$$

$$a = \bar{y} - b\bar{x}$$

Donde: n es el número de observaciones, “x” es la media de todos los valores de “x”; y “y” es la media de todos los valores de “y”.

En nuestro ejemplo encontramos, utilizando una calculadora apropiada:

<i>Ingreso familia en \$/año ($X_i$)</i>	<i>Peso en kg (Y_i)</i>	$X_i Y_i$	X_i^2	Y_i^2
130	15,5	2.015,00	16.900	240,25
200	19,8	3.960,00	40.000	392,04
345	21,5	7.414,50	119.025	462,25
245	16,8	4.116,00	60.025	282,24
155	12,6	1.953,00	24.025	158,76
300	16,6	4.980,00	90.000	275,56
360	18,1	6.516,00	129.600	327,61
105	18,7	1.963,50	11.025	349,69
80	13,1	1.048,00	6.400	171,61
275	20,1	5.527,50	75.625	404,01
225	18,1	4.072,50	50.625	327,61
95	17,4	1.653,00	9.025	302,76
130	17,9	2.327,00	16.900	320,41
330	17,0	5.610,00	108.900	289,00
295	18,7	5.516,50	87.025	349,69
170	16,0	2.720,00	28.900	256,00
250	18,2	4.550,00	62.500	331,24
355	16,4	5.822,00	126.025	268,96
220	15,4	3.388,00	48.400	237,16
175	17,6	3.080,00	30.625	309,76
Total 4.440	345,5	7.8235,50	1.141.550	6056,61

$$\bar{x} = 222, \bar{y} = 17,275 \quad \Sigma xy = 78.235,5$$

$$\Sigma x = 4440, \quad \Sigma y = 345,5$$

$$\Sigma x^2 = 1141550, \quad \Sigma y^2 = 6056,6$$

$$(\Sigma x)^2 = 19713600, \quad (\Sigma y)^2 = 119370,25$$

$$b = \frac{78235,5 - (4440)(345,5) / 20}{1141550 - (19713600) / 2} = 0,00984$$

$$a = 17,275 - 0,00984(222) = 15,09$$

Así, la ecuación de nuestra línea ajustada es

$$y = 15,09 + 0,00984x$$

Anexo 31.2 Cálculo del coeficiente de correlación de Pierson

La fórmula para calcular el coeficiente de correlación es la siguiente:

$$r = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\sqrt{(\Sigma(x - \bar{x})^2 \Sigma(y - \bar{y})^2)}} = \frac{\Sigma xy - (\Sigma x)(\Sigma y) / n}{\sqrt{(\Sigma x^2 - (\Sigma x)^2 / n)(\Sigma y^2 - (\Sigma y)^2 / n)}}$$

En nuestro ejemplo del peso y el ingreso familiar, esto significaría:

$$r = \frac{78235,5 - (4440)(345,5) / 20}{\sqrt{(1141550 - (19713600) / 20)(6056,6 - (119370,25) / 20)}}$$

$$r = 0,414$$

(Para su interpretación, véase el inciso IV de este módulo.)

**Módulo 31 MEDICIÓN DE LAS ASOCIACIONES ENTRE
LAS VARIABLES; REGRESIÓN Y CORRELACIÓN**

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
3 horas	Trabajo en equipo

Introducción y discusión

- Este módulo debe presentarse sólo si al menos uno de los equipos de investigación lo necesita para analizar sus datos.
- Es preferible presentar la sección de la correlación antes de la sección de la regresión (por ejemplo, mostrando los dos primeros diagramas de dispersión de la sección IV después de la *ilustración 31.1*, y luego continuar con la sección III), ya que de este modo es más probable que los participantes puedan entenderla y usarla en futuros proyectos.

Los facilitadores deben cerciorarse de explicar *cuándo* emplear el análisis de correlación y cuándo el de regresión.

Módulo 32

REDACCIÓN DEL INFORME DE INVESTIGACIÓN

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 32 REDACCIÓN DEL INFORME DE INVESTIGACIÓN

OBJETIVOS

Después de esta sesión, usted debe ser capaz de:

1. *Enumerar* los principales componentes de un informe de investigación.
2. *Hacer* un esquema de su informe.
3. *Escribir* borradores por etapas.
4. *Revisar* que el borrador final esté completo, que no tenga traslapes y que esté escrito en un estilo claro y fluido.
5. *Esbozar* recomendaciones para la acción con base en sus hallazgos.

I. Pasos de la preparación del informe: consideraciones preliminares.

II. Redacción del informe de investigación.

1. Introducción
2. Principales componentes de un informe de investigación
3. Estilo y presentación
4. Puntos débiles frecuentes en la redacción
5. Revisión y terminación del texto

I. PREPARACIÓN DEL INFORME DE INVESTIGACIÓN: CONSIDERACIONES PRELIMINARES

¿Quiénes van a leer su informe de investigación? ¿Cómo lo van a leer?

Los estudios de ISS tienen diferentes lectores: funcionarios de la salud, investigadores y miembros de la comunidad interesados en el tema. Cada uno de estos grupos va a leer su informe desde una perspectiva diferente.

Los funcionarios de la salud y los miembros de la comunidad van a preguntarse:

- *¿Cómo ayuda esta “nueva información” a mejorar la salud de la comunidad?* (es decir, ¿cuál es el problema y en qué medida lo resuelve o lo atenúa la información?)

Los investigadores, por su parte, querrán saber:

- *¿Puedo “creer” en estos hallazgos?* (es decir, ¿los hallazgos son válidos y confiables?). El diseño de la investigación, el muestreo, los métodos de recolección de datos y el análisis de datos deben sustentar su validez y su fiabilidad.

Así, los informes de las ISS deben satisfacer las necesidades de los *funcionarios de la salud*, de los *investigadores* y del (los) *grupo(s) objeto* de la investigación.

II. REDACCIÓN DEL INFORME DE INVESTIGACIÓN

1. Introducción

Los informes de ISS deben:

- tener una clara estructura lógica
- ir al grano
- emplear un lenguaje sencillo y tener una presentación agradable

Así como un arquitecto que diseña una casa elabora un plan, usted debe hacer, antes que nada, un ESQUEMA de su informe. Este esquema debe tener una cabeza, un cuerpo y un final. La cabeza consiste en una descripción del problema en su contexto (país y área de investigación), los objetivos del estudio y el método de investigación aplicado. Esta sección no debe conformar más de una cuarta parte del informe, de otro modo éste se hace demasiado pesado. El

cuerpo constituye la mayor parte de su informe: va a contener los hallazgos de la investigación. El final, por último, consiste en comentarios sobre los datos, las conclusiones y las recomendaciones.

Una vez terminado el esquema, haga atractivo el informe, y fácil de leer y consultar, con una carátula creativa, un prefacio con los agradecimientos, un índice, tal vez la lista de las tablas, otra de las ilustraciones y otra más de las abreviaturas. Por supuesto, habrá que agregar la bibliografía que utilizó para su estudio, y los anexos (incluyendo, por lo menos, sus herramientas de recolección de datos).

Antes de empezar a escribir, es esencial agrupar y revisar los datos que analizó *por objetivo*. Compruebe que realmente haya procesado y analizado todos los datos como lo planificó en el trabajo en equipo del *módulo 21*.

Esboce sus conclusiones más importantes y relaciónelas con la bibliografía que leyó. Una vez más, puede sentirse inspirado para regresar a sus datos en bruto para refinar su análisis o buscar bibliografía adicional que le permita responder preguntas que el análisis de los datos pueda suscitar.

Reúna las principales conclusiones y las tablas o las citas de los datos cualitativos relacionados con cada objetivo. Ahora está usted listo para redactar el informe.

2. Principales componentes del informe de investigación

El informe de investigación debe contener los siguientes componentes:

TÍTULO y CARÁTULA

RESUMEN DEL DISEÑO DEL ESTUDIO, HALLAZGOS Y RECOMENDACIONES

AGRADECIMIENTOS

ÍNDICE

Lista de tablas e ilustraciones (opcional)

Lista de abreviaturas (opcional)

1. INTRODUCCIÓN (planteamiento del problema en su contexto local, incluyendo la bibliografía pertinente)
2. OBJETIVOS
3. MÉTODO DE INVESTIGACIÓN
4. HALLAZGOS DE LA INVESTIGACIÓN
5. COMENTARIOS

6. CONCLUSIONES Y RECOMENDACIONES

REFERENCIAS

ANEXOS (instrumentos de la recolección de datos, tablas)

Los hallazgos, los comentarios, las conclusiones y las recomendaciones constituyen la parte sustancial del informe, que debe redactarse desde cero.

Por lo tanto, le recomendamos *comenzar con los hallazgos, los comentarios y las conclusiones*. No obstante, vamos a comentar brevemente cada uno de los componentes mencionados en la secuencia en la que van a aparecer, finalmente, en el informe.

- **Carátula**

La carátula debe contener el título, los nombres de los autores con sus títulos y nombramientos, la institución que publica el informe (por ejemplo, Ministerio de Salud, ISS) y el mes y el año de publicación. El título podría consistir en una declaración o una pregunta desafiante, seguida de un subtítulo informativo que cubra el contenido del estudio e indique la región donde el estudio se llevó a cabo. (Véase el módulo 6).

- **Resumen**

El resumen debe redactarse sólo *después* de haber terminado el primer –o incluso del segundo– borrador del informe. Éste debe contener:

- una muy breve descripción del problema (QUÉ se estudió)
- los principales objetivos (POR QUÉ era necesario este estudio)
- el lugar del estudio (DÓNDE)
- el tipo de estudio y los métodos empleados (CÓMO)
- los principales hallazgos y conclusiones
- las principales (o todas) las recomendaciones.

El resumen es la primera parte del estudio y muy probablemente la única que lean los responsables de la toma de decisiones para la salud, puesto que están llenos de ocupaciones. Por lo tanto, su redacción implica una profunda reflexión y toma mucho tiempo. Es posible que deba escribir varios borradores y comentar cada uno con el equipo de investigación en su conjunto.

Si lo considera necesario puede redactar un resumen distinto para cada grupo con que haya colaborado durante la elaboración y la ejecución de su propuesta de investigación. Por ejemplo, uno para los responsables de la toma de decisiones y los funcionarios de la salud, otro para el personal de la salud de niveles inferiores, otro para los miembros de la comunidad o para el público en general (el periódico, la televisión). En una etapa posterior puede escribir artículos en revistas científicas. (Vea el módulo 33).

- **Agradecimientos**

Es una buena costumbre agradecer a quienes lo apoyaron técnica o económicamente en el diseño y la implementación de su estudio. También pueden recibir un reconocimiento su jefe inmediato, quien le permitió invertir tiempo en el estudio, y las personas que respondieron la encuesta. Los agradecimientos suelen ir en la página siguiente a la del título o al final del informe, antes de la bibliografía.

- **Índice**

El índice es esencial, ofrece al lector un panorama general de las principales secciones de su informe, con referencias a las páginas, de modo que puede recorrer el informe en un orden diferente o saltarse algunas secciones.

- **Lista de las tablas y las ilustraciones**

Si usted tiene muchas tablas o ilustraciones es útil presentarlas también en una lista, en un formato del tipo del “índice,” con números de página.

- **Lista de abreviaturas (opcional)**

Si hace usted uso de abreviaturas, siglas y acrónimos en la informe, estos deben citarse con todas sus letras en el texto la primera vez que se mencionan. Si son muchos, también deben enumerarse en orden alfabético. La lista puede colocarse antes del primer capítulo del informe.

El índice y las listas de las tablas, las ilustraciones y las abreviaturas deben prepararse al final, porque sólo entonces cuenta con el número definitivo de página de cada capítulo y sub-sección. Es también entonces cuando puede terminar de numerar las tablas y las ilustraciones e incluir todas las abreviaturas.

Capítulo 1: Introducción

La introducción es una sección relativamente fácil de redactar, que es mejor escribir después de haber hecho un primer borrador de los hallazgos. Sin duda,

debe contener algunos datos contextuales (del medio ambiente, administrativos, económicos y sociales) pertinentes sobre el país, el estado de salud de la población y de los servicios de salud relacionados con el problema estudiado. Puede añadir poco a poco esta información en las secciones correspondientes en su propuesta de investigación, incluyendo la bibliografía pertinente, y utilizar todo esto en su informe.

Enseguida debe ir el planteamiento del problema, de nuevo revisado a partir de su propuesta de investigación, con comentarios adicionales y la bibliografía recopilada durante la elaboración del estudio. Debe incluir un párrafo sobre sus expectativas en torno al estudio.

En la introducción al planteamiento del problema puede citar la bibliografía internacional, si el problema de su estudio es de interés internacional. De otro modo, la bibliografía relevante de distintos países puede seguir al planteamiento del problema como una revisión bibliográfica independiente. También puede presentar los conceptos o modelos teóricos que haya utilizado en el análisis de los datos, en una sección por separado después del planteamiento del problema.

Capítulo 2: Objetivos

Deben incluirse los objetivos generales y los específicos tal como se indicaron en la propuesta de investigación. De ser necesario, puede ajustarlos ligeramente en función del estilo y la secuencia. Sin embargo, no debe cambiar su naturaleza básica. Si usted no logró cumplir algún objetivo, debe asentarlos en la sección donde describe el método de investigación y en la sección donde comenta los hallazgos. Los objetivos constituyen el CORAZÓN de su estudio. Determinaron el método que eligió, y determinarán la manera en que va a estructurar el informe de sus hallazgos.

Capítulo 3: Método de investigación

El método de investigación empleado en la recolección de datos debe describirse pormenorizadamente. Esta sección debe incluir una descripción de:

- el tipo de estudio
- los principales temas o variables del estudio (puede anexarse una lista pormenorizada de las variables sobre las que se recogieron datos)
- la población de estudio, el método de muestreo y el tamaño de cada muestra
- las técnicas de recolección de datos utilizadas para cada población de estudio

- cómo se recogieron los datos y por quién
- los procedimientos empleados para el análisis de los datos, incluyendo las pruebas estadísticas (si cabe).

Si se alejó usted del diseño original del estudio presentado en su propuesta de investigación, debe explicar en qué medida lo hizo y por qué. Indique las consecuencias de esta desviación para alcanzar determinados objetivos del estudio. Si la calidad de algunos de los datos es pobre y ello puede dar lugar a un sesgo, también debe describirlo bajo el encabezado “limitaciones del estudio”.

Capítulo 4: Hallazgos de la investigación

La presentación sistemática de sus hallazgos en relación con los objetivos de la investigación es la parte crucial de su informe.

La descripción de los resultados debe ofrecer una buena combinación o triangulación de los datos de los componentes cualitativos y cuantitativos del estudio.

Puede usted presentar sus hallazgos de dos diferentes maneras:

(1) En una presentación integrada de todos los datos por objetivo

Puesto que enlistó sus datos por objetivo (véase el módulo 21), esto debería ser fácil.

Por ejemplo, en un estudio de los factores asociados con la alta mortalidad materna, las entrevistas con los familiares de las madres que murieron en el parto revelaron que 90% de ellas habían recibido atención prenatal. Sólo 45% había dado a luz alguna vez en instalaciones de salud y sólo 25% lo había hecho así durante el parto que resultó mortal. El análisis de las tarjetas de atención prenatal con el personal local de salud indicó que alrededor de 35% de las madres que visitaron el centro de atención prenatal en su último embarazo, había dado a luz en un hospital o un centro de salud. Dicho análisis confirma, por consiguiente, el escaso uso –en general– de la atención del parto en un centro de salud, aunque este uso parece ser ligeramente superior al de las madres que murieron. Las entrevistas con los familiares de las madres que fallecieron pusieron de manifiesto que la distancia y el costo eran los principales factores para no asistir a las instalaciones de salud a ser atendidas en el parto. Los GD con las madres también revelaron la preocupación por la falta de privacidad en los centros de salud, el descuido del personal de salud y la corta edad de las parteras. Una mujer (40 años) declaró: *En ese hospital las parteras son muy jóvenes, ellas mismas aún no han dado a luz. ¿Cómo puedo ser atendida por mi hija?, ¿cómo puedo dejar que me vea en ese estado?*

Estas presentaciones integradas van a constituir una compilación de tablas, gráficos, interpretación de las narraciones y citas ilustrativas de las entrevistas a profundidad o los GD.

(2) En una presentación de los datos por instrumento de investigación

A veces es más fácil analizar los datos por instrumento e integrar los hallazgos en la sección de los comentarios. El análisis independiente está indicado para los objetivos que cubren distintas poblaciones de estudio, cada una mediante instrumentos particulares.

Por ejemplo, en el estudio de los motivos del incumplimiento de los pacientes con tratamiento para la tuberculosis presentado en el *módulo 4*, se emplearon cuatro diferentes instrumentos de investigación. Mediante el análisis de las tarjetas clínicas de una cohorte de pacientes de tuberculosis, se pudo definir la proporción de pacientes irregulares y la de los que no asistieron para recibir su tratamiento (objetivo 1).

Los GD con diferentes grupos de miembros de la comunidad brindaron datos sobre la manera en que la comunidad percibía la enfermedad (objetivo 2). Las entrevistas con los pacientes (regulares, irregulares y los que no asistieron) arrojan luz sobre las diferentes razones por las que no se siguió el tratamiento (objetivos 3 y 4). Las entrevistas con el personal de salud revelaron las deficiencias de los servicios que podrían estar contribuyendo a que los pacientes no siguieran el tratamiento (objetivo 5). La discusión reveló parte de los datos que debían integrarse (por ejemplo, algunos miembros de la comunidad –y los propios pacientes– también tenían una opinión de las deficiencias o las fortalezas de los servicios, mientras que el personal y los miembros de la comunidad complementaron los datos de los pacientes sobre las razones económicas y socio-culturales de la deserción).

La lista de los datos por objetivo le ayudará a decidir cómo organizar su presentación en el informe. La decisión de dónde poner qué puede hacerse mejor una vez que se han procesado y analizado todos los datos, y antes de comenzar a escribir.

Cuando haya analizado todos los datos, haga un ESQUEMA pormenorizado para presentar los hallazgos. Esto le va a ayudar a decidir cómo organizar los datos, y es una *condición indispensable para la división óptima de las tareas entre los miembros del grupo en el proceso de la redacción*.

En esta etapa puede preparar también un esquema del informe completo, tomando como punto de partida los elementos principales de un informe de investigación (p. 5).

Un esquema debe contener:

- Los títulos de las principales secciones del informe
- Los títulos de las sub-secciones
- Los puntos que deben tratarse en cada sección
- Listas de las tablas, las ilustraciones y las citas para ilustrar cada sección

Es de esperar que el esquema de la sección de los hallazgos sea el más elaborado.

La primera parte de la sección de los hallazgos suele ser la descripción de la población de estudio. Cuando haya estudiado diversas poblaciones, debe describir las brevemente antes de presentar los datos correspondientes a los informantes.

Entonces, dependiendo del diseño del estudio, puede proporcionar más información del problema que estudió (tamaño, distribución, características). En un estudio analítico, se comenta posteriormente el grado en que las distintas variables influyen en el problema.

Por ejemplo: En un estudio sobre desnutrición, el capítulo “Conclusiones” puede verse así:

Capítulo 4: Hallazgos

4.1 Descripción de la muestra

(Por ejemplo, ubicación, edad, estado civil, educación, estatus socioeconómico de la madre; edad y sexo de los niños pesados y medidos por área de investigación).

4.2 Amplitud y variación estacional de la desnutrición en el distrito X

4.3 Posibles causas de la desnutrición

4.3.1 Disponibilidad limitada de alimentos

4.3.2 Uso deficiente de los alimentos disponibles

4.3.3 Prevalencia de enfermedades transmisibles

4.3.4 Acceso limitado a los servicios materno infantiles y deficiencias en dichos servicios y en los servicios de nutrición

Este sistema de numeración es flexible y puede ampliarse según las necesidades con más títulos o subtítulos. Permite mantener un panorama general del proceso, lo que es útil cuando diferentes miembros del grupo trabajan al mismo tiempo en distintas secciones del informe.

Si sus hallazgos son tan elaborados que deba usted tener sub-sub-sub-secciones de 4 o 5 dígitos, puede optar por dividir las conclusiones en varios capítulos. Considere dejar fuera algunas de las numeraciones de las sub-secciones, si queda claro a qué sección mayor pertenecen. Sin embargo, conserve la numeración completa hasta que escriba el borrador definitivo, ya que ésta le va a ayudar a mantener su informe en orden aun cuando diferentes miembros del grupo trabajen al mismo tiempo en distintas secciones.

Las TABLAS y las ILUSTRACIONES distribuidas a lo largo del texto deben estar numeradas y tener títulos claros. Es aconsejable utilizar primero el número de la sección a la que pertenece la tabla. En el último borrador puede numerar las tablas y las ilustraciones en una secuencia.

Incluya sólo las tablas y las ilustraciones que presentan los hallazgos más importantes, que necesiten ser comentadas de manera más elaborada a lo largo del texto. Otras pueden presentarse en los anexos o, si no revelan puntos interesantes, pueden omitirse.

Tenga en cuenta que no es necesario describir detalladamente cualquier tabla sólo porque la incluye el informe. Presente sólo las conclusiones más importantes.

Nota:

El primer borrador de sus hallazgos nunca es el definitivo. Por lo tanto, puede concentrarse en el contenido más que en el estilo. No obstante, es aconsejable estructurar el texto en párrafos desde un principio y formular cada oración de manera clara y precisa.

Capítulo 5: Comentarios

Ahora puede comentar los resultados por objetivo o por grupos de variables o temas relacionados, lo que deriva en conclusiones y posibles recomendaciones. Los comentarios pueden incluir hallazgos de otros estudios relacionados que apoyen o contradigan el suyo.

Capítulo 6: Conclusiones y recomendaciones

Las conclusiones y las recomendaciones se derivan lógicamente de la discusión de los hallazgos. Las conclusiones pueden ser breves puesto que ya se comentaron profusamente en el capítulo 5. Dado que los comentarios siguen la secuencia en que se presentan los hallazgos (que dependen, a su vez, de los objetivos), las conclusiones deben seguir el mismo orden.

La lectura de una persona ajena al tema se facilita si también las recomendaciones se colocan más o menos en la misma secuencia de las conclusiones. Sin embargo, las recomendaciones pueden resumirse al mismo tiempo, de acuerdo con los grupos a los que se dirigen, por ejemplo:

- las personas responsables de la toma de decisiones
- funcionarios de la salud o relacionados con la salud, del distrito o de niveles inferiores
- personal de la salud o relacionado con la salud, que pueda instrumentar las actividades
- los clientes potenciales
- la comunidad en general

Recuerde que los grupos orientados a la acción son los más interesados en esta sección.

Al formular las recomendaciones, use los hallazgos de su estudio, y también la información de apoyo de otras fuentes. Las recomendaciones deben considerar las características locales del sistema de salud, las limitaciones, la factibilidad y la utilidad de las soluciones propuestas. Éstas deben comentarse con todos los interesados antes de su redacción final.

Si sus recomendaciones son breves (de aproximadamente una página), puede incluirlas todas en el resumen y omitirlas como una sección aparte en el capítulo 6 para evitar su repetición.

Referencias

Puede usted numerar las referencias hechas a lo largo del texto en el orden en que aparecen en el informe y citarlas luego en el mismo orden en la lista de referencias (sistema de Vancouver), o emplear el sistema de Harvard: cita entre paréntesis en el texto el nombre de los autores de la referencia seguido de la fecha de publicación y el número de página; por ejemplo: (Shan, 2000: 84). En la bibliografía, disponga las publicaciones en orden alfabético por el apellido del autor principal. (Vea el módulo 5.)

Puede optar por cualquier sistema, siempre y cuando lo use consistentemente en todo el informe.

Anexos o apéndices

Los anexos deben contener cualquier información adicional necesaria para permitir a los profesionales seguir sus procedimientos de investigación y análisis de los datos.

Toda aquella información que es útil para determinadas categorías de lectores, pero no es del interés del lector medio también puede incluirse en anexos.

La siguiente información puede presentarse en los anexos:

- las tablas que se mencionan en el texto, pero no se incluyen a fin de controlar la extensión del informe
- listas de los hospitales, distritos, poblados, etc. que participaron en el estudio;
- los cuestionarios o las listas de verificación utilizadas en la recolección de datos.

Nota:

Nunca empiece a escribir sin un esquema. Antes de procesar el texto, todas las secciones deben tener título y número conforme al esquema. Tenga a la vista el esquema (en la pared) para que todos se percaten de inmediato de cualquier adición o cambio, o de los progresos realizados.

Prepare un primer borrador de su informe a doble espacio y con márgenes amplios, de modo que pueda usted hacer fácilmente comentarios y correcciones en el texto.

Tenga varias copias del primer borrador para trabajar en ellas y una copia para insertar los cambios finales para su revisión.

3. Estilo y presentación

(1) Estilo de la redacción

Recuerde que su lector:

- Tiene poco tiempo
- Tiene muchos otros asuntos urgentes que demandan su interés y su atención

- Probablemente no esté al tanto de la “jerga de la investigación”

Por consiguiente, considere las siguientes normas:

- Simplifique. Límitese a lo esencial
- Justifique. No haga ninguna declaración que no esté fundada en hechos y datos
- • Cuantifique cuando cuente con los datos necesarios. Evite expresiones imprecisas como “grande” o “pequeño”; escriba, en cambio, “50%”, “uno de cada tres”
- Sea preciso y concreto al formular los hallazgos
- Informe, no impresione. Evite la exageración
- Construya oraciones cortas.
- Utilice con moderación los adverbios y los adjetivos
- Sea consistente en el uso de los tiempos verbales (pasado o presente). De ser posible, evite la voz pasiva, ya que genera ambivalencia (por ejemplo, “los pacientes fueron entrevistados” genera incertidumbre con respecto a quién los entrevistó), y su uso reiterado hace aburrida la lectura
- Procure que su presentación sea lógica y sistemática.

(2) Presentación del informe

Una buena presentación física es importante, ya que contribuye a que su informe:

- dé una buena primera impresión
- aliente a sus lectores
- dé una idea de cómo está organizado el material para que el lector decida rápidamente qué leer primero.

Preste especial atención a que:

- La página del título tenga un diseño atractivo y el índice sea claro
- Los márgenes y el espaciado sean consistentes
- Haya consistencia en los títulos y subtítulos; por ejemplo, en el tamaño de fuente: 16 o 18 en negrita para los títulos de los capítulos; 14 en negrita para los títulos de las secciones principales; 12 en negrita para los títulos de las sub-secciones, etc.

- La impresión y el fotocopiado tengan una buena calidad. Corrija los borradores cuidadosamente; revise la ortografía; solicite a otros miembros del equipo, a su facilitador y, de ser posible, a personas ajenas al proyecto, una lectura crítica para llegar a una mayor claridad
- Las ilustraciones y las tablas estén numeradas, las tablas tengan títulos claros, y las columnas y las filas tengan encabezados igualmente claros
- Las citas y las referencias sean precisas y coherentes

4. Deficiencias comunes en la escritura

La escritura siempre representa un desafío que requiere valor. Comenzar suele ser lo más difícil. No tema cometer errores, de lo contrario ¡no va a empezar nunca! No obstante, es bueno ser conscientes de las trampas más comunes, que usted puede evitar.

Una deficiencia casi universal de los autores primerizos de un informe es la *omisión de lo evidente*. Dificilmente la descripción del país o del área contiene los datos suficientes que permitan a los forasteros seguir la presentación de los hallazgos y los comentarios sin problemas. Por otra parte, contiene algunos datos (como, por ejemplo, la ubicación geográfica exacta en el mapamundi) que bien podrían quedar fuera.

Las *descripciones interminables* sin una interpretación representan otra trampa. Las tablas necesitan conclusiones, y no la presentación pormenorizada de todos los números o porcentajes de las celdas que los lectores pueden ver por sí solos. El capítulo de los comentarios, en particular, necesita datos comparados que destaquen los resultados imprevistos, su propia opinión —o la de otros— de los problemas descubiertos, la ponderación de los pros y los contras de las posibles soluciones. Sin embargo, con demasiada frecuencia los comentarios consisten tan solo en un resumen insípido de los hallazgos.

La *omisión de los datos cualitativos* es también bastante común. No obstante, la ilustración de sus hallazgos y conclusiones con citas de los informantes agiliza el informe. Hay también un valor científico en permitir que el lector saque sus propias conclusiones a partir de los datos que usted presenta. (¡Dando por hecho que su presentación no está sesgada!)

A veces los datos cualitativos (por ejemplo, las preguntas de opinión abiertas) simplemente se codifican y se cuentan como datos cuantitativos, *sin interpretarlos*, siendo que podrían brindar interesantes ilustraciones de las razones del comportamiento o actitudes de los informantes. Presentar así los datos es grave; debe corregirse.

5. Revisión y conclusión del texto

Una vez terminado un primer borrador de los hallazgos, los comentarios y las conclusiones, todos los miembros del grupo de trabajo y los facilitadores deben hacer una lectura crítica y aportar comentarios.

Al leer el borrador, pregúntese:

- ¿Todos los hallazgos importantes están presentes?
- ¿Las conclusiones se desprenden lógicamente de los hallazgos? Si algunos de los hallazgos se contradicen entre ellos, ¿hay una explicación?, ¿ya se comentó y se explicó? ¿De haber deficiencias metodológicas, ya se señalaron?
- ¿En el borrador hay traslapes que deben eliminarse?
- ¿Es posible condensar el contenido? Por lo general un texto gana cuando se abrevia. Algunas partes menos relevantes para la acción pueden incluirse en los anexos. Revise si los párrafos descriptivos pueden abreviarse y *presentarse o terminarse con una oración concluyente*.
- ¿Los datos del texto concuerdan con los datos de las tablas? ¿Todas las tablas son consistentes (con el mismo número de informantes por variable), tienen títulos y encabezados claros y la numeración tiene una secuencia?
- ¿La secuencia de las secciones y sub-secciones es lógica y coherente? ¿La conexión entre los párrafos y las secciones es fluida? ¿La formulación de los hallazgos y las conclusiones es clara y precisa?

Los autores de las distintas secciones pueden preparar un segundo borrador, considerando las observaciones. O también se puede nombrar a dos participantes como editores para que redacten la versión completa.

Mientras tanto, otros miembros del grupo pueden (re)escribir la sección introductoria (INTRODUCCIÓN, OBJETIVOS y MÉTODO DE INVESTIGACIÓN, ajustándola a su propuesta original).

En este punto puede redactarse un primer borrador del resumen (véase la página 5 de este módulo).

Terminación del informe de investigación

Es recomendable que otros grupos y facilitadores revisen el segundo borrador considerando los puntos mencionados en la sección anterior. Después, debe prepararse la versión final del informe. Esta vez ponga especial cuidado en la presentación: la estructura, el estilo y la consistencia ortográfica (¡use el corrector ortográfico!).

Use consistentemente los tiempos verbales. Las descripciones de la situación en el campo pueden expresarse en pasado (por ejemplo, “cinco familias tenían menos de una hectárea de tierra”). Las conclusiones extraídas de los datos se expresan, por lo general, en presente (por ejemplo, “Los tabúes alimentarios difícilmente afectan el estatus nutricional de los niños pequeños”).

Nota:

Para hacer una comprobación final de la legibilidad del informe, hojéelo y lea la primera oración de cada párrafo. Si le da una impresión clara de la organización y los resultados de su estudio, puede concluir que lo hizo lo mejor que se podía.

TRABAJO EN EQUIPO

1. *Haga un esquema de su informe* en una hoja de rota-folio, después de revisar sus objetivos, sus fuentes de información y los resultados del análisis de datos. Haga una propuesta de numeración de las secciones y las sub-secciones. Pegue el esquema en la pared en un lugar visible. Deje el espacio suficiente entre los renglones para poder agregar, por ejemplo, sub-secciones, y para anotar cualquier cambio.
2. *Empiece a redactar por el capítulo de los hallazgos.* Decida con su facilitador si va a presentar la interpretación de los datos por variable, por objetivo o por población de estudio. Si al principio no está seguro de qué método podría funcionar mejor, registre sus hallazgos e interpretaciones en función de la población de estudio. En el segundo borrador puede reorganizar y abreviar la presentación. *Divida las tareas de la redacción entre sub-grupos de una o dos personas.*
3. *Comente sus hallazgos relacionándolos unos con otros, con los objetivos y con otros textos bibliográficos, y redacte el capítulo de los Comentarios.* Enseguida haga una lista de las principales conclusiones en relación con posibles recomendaciones.
4. *Elabore, al mismo tiempo, los capítulos introductorios* (contexto y planteamiento del problema –incluya nueva bibliografía–, objetivos y método de investigación), adapte lo que preparó antes para la propuesta de investigación.
5. *Redacte, finalmente, el resumen* siguiendo el esquema presentado anteriormente en este módulo. Tómese por lo menos medio día de trabajo sistemático.
6. *Lleve un registro de los avances* en la redacción y la transcripción, tomando notas en el rota-folio con el esquema de su informe.
7. *Revise el primer borrador con todo el grupo; descarte* lagunas, redundancias, etc., antes de preparar el segundo borrador. Pídale al facilitador de otro grupo que lea el borrador del informe completo antes de darlo por terminado.

Referencias

- Gibaldi J (1995). *MLA Handbook for Writers of Research Papers*, New York: Modern Language Association of America.
- Jen Tsi Yang et al. (1996). *An outline of Scientific Writing: For Researchers with English As a Foreign Language*, Singapore: World Scientific Publishing, www.amazon.com/exec/obidos (through internet, October 2000)
- Lindsay D (1996). *Guide to Scientific Writing*, Australia: Addison & Wesley (paperback), www.amazon.com/exec/obidos (through internet, October 2000).

Notas del instructor

Módulo 32 REDACCIÓN DEL INFORME DE INVESTIGACIÓN

Tiempo estimado y recursos didácticos

1 hora	Introducción y discusión
Varios días	Trabajo en equipo

Introducción y discusión

- Ponga el esquema del informe de investigación en una lámina de retroproyector y coméntelo punto por punto. Subraye la prioridad de los hallazgos, comentarios, conclusiones y recomendaciones.
- Tome un ejemplo de alguno los grupos cuando presente un posible esquema del capítulo de los *Hallazgos*, con títulos y subtítulos apropiados.
- Explíqueles a los participantes el sistema de numeración y la diferenciación de los tamaños de la fuente, la consistencia en la presentación de los títulos y los subtítulos, de modo que más adelante pueda usted utilizar el ejemplo para ilustrar el diseño adecuado.
- Pídales a los participantes –antes de darles ninguna directriz–, que propongan los criterios con que juzgarían su primer borrador.
- Para explicar cómo adaptar el planteamiento del problema, los objetivos y el método de investigación para el informe final, utilice ejemplos de las propuestas de investigación preparadas por los diversos grupos.

Si considera que puede no estar suficientemente claro, insista en la necesidad de sustituir el tiempo futuro usado en la propuesta de investigación por los tiempos presente y pasado.

Trabajo en equipo

- Cerciórese de que todos los grupos hagan primero un esquema de sus informes, utilizando como punto de partida el esquema presentado en este módulo. Pídales que lo cuelguen de modo que todos los miembros del equipo lo puedan ver.
- Las secciones sobre los hallazgos, los comentarios y las conclusiones y recomendaciones toman más tiempo. Quizás para algunos de los grupos la presentación de estas secciones pueda funcionar mejor en un orden diferente. Hágales saber que pueden utilizar el esquema de presentación que les parezca más apropiado para sus datos, pero permita que los facilitadores les sugieran cómo estructurar de la manera más lógica esta parte de su presentación.
- Quizás sea mejor comenzar la redacción por los hallazgos y las conclusiones. Sólo una vez que esté listo un borrador razonablemente estructurado pueden hacerse recomendaciones a los miembros del equipo de investigación para la reescritura de los capítulos introductorios.
- Todos los miembros del grupo deben redactar alguna sección del informe. Si es necesario, arme grupos de redacción por parejas. Si algunos de los participantes nunca han escrito nada, puede necesitar un apoyo intensivo. Déjelos escribir varios párrafos y luego reescribalos con ellos, para darles un ejemplo.
- Si los grupos no tienen experiencia en la redacción de informes, van a necesitar una orientación explícita sobre los puntos en los que deben poner especial atención al revisar su primer borrador, incluyendo el esquema básico del informe.
- A la hora de redactar los hallazgos y comentarios, proponga una lluvia de ideas de posibles recomendaciones. Regístrelas de inmediato (de preferencia en un rota-folio) para que puedan usarse cuando se formulen las recomendaciones.
- Todos los facilitadores deben comentar la sección del resumen que cada grupo preparó. Cada facilitador también debe leer y comentar al menos un borrador completo del informe de algún otro grupo antes de que éste pueda darse por terminado.

Módulo 33

DIFUSIÓN, COMUNICACIÓN Y USO DE LOS HALLAZGOS DE LA INVESTIGACIÓN

Pasos del análisis de los datos y redacción del informe

Preguntas que deben hacerse	Pasos que deben darse*	Elementos importantes en cada paso
¿Qué datos se han recogido para cada uno de los objetivos de la investigación? ¿Los datos están completos y son precisos?	Prepare los datos para su análisis	<ul style="list-style-type: none">– Revise la experiencia de campo– Haga un inventario de los datos de cada uno de los objetivos/población de estudio– Clasifique los datos y revise su calidad– Revise las salidas de la computadora (21)
¿Cómo lucen los datos? ¿Cómo sintetizarlos para facilitar su análisis?	Resuma los datos y describa las variables / identifique nuevas variables	<ul style="list-style-type: none">– Tablas de frecuencia, gráficos, significados, proporciones, tabulaciones cruzadas descriptivas (<i>datos cuantitativos</i>) (22, 24); Codificación, listas, síntesis de los datos en hojas de compilación, matrices, diagramas de flujo, diagramas y narraciones (<i>datos cualitativos</i>) (23)
¿Cómo determinar las asociaciones entre las variables?	Analice las asociaciones	<ul style="list-style-type: none">– Tablas cruzadas analíticas (24)– Medidas de la asociación basada en el riesgo (25)– Manejo de las variables de confusión (26)
	Prepare el análisis estadístico	<ul style="list-style-type: none">– Medidas de la dispersión, la distribución normal y las variaciones de la muestra (27)
¿Medimos las diferencias o las asociaciones entre las variables?	Determine los tipos de análisis estadísticos	<ul style="list-style-type: none">– Selección de las pruebas de significación (28)
¿Cómo determinar las diferencias entre los grupos?	Analice las observaciones apareadas y sin aparear	<ul style="list-style-type: none">– Pruebas t y chi cuadrado (χ^2) (29)** prueba t para datos apareados**prueba de chi cuadrado (χ^2) de McNemar (30)
¿Cómo determinar las asociaciones entre las variables numéricas?	Establezca medidas de asociación	<ul style="list-style-type: none">** Diagrama** Línea de regresión** Coeficiente de correlación (31)
¿Cómo redactar el informe?	Escriba el informe y formule recomendaciones	<ul style="list-style-type: none">– Prepare un bosquejo del informe– Presente e interprete los datos– Formule y reformule– Comente y sintetice las conclusiones– Formule las recomendaciones (32)
¿Cómo comunicar, difundir y utilizar los hallazgos y las recomendaciones?	Presente la síntesis y la formulación de las recomendaciones para implementarlas	<ul style="list-style-type: none">– Comente con todos los involucrados las síntesis y el plan para la implementación (33)

* Estos pasos no tienen necesariamente la secuencia que se muestra en el diagrama. *La secuencia puede ajustarse de acuerdo con las necesidades de los equipos de investigación.*

** Estos elementos son opcionales y pueden omitirse de no ser relevantes para los equipos de investigación.

Módulo 33 DIFUSIÓN, COMUNICACIÓN Y USO DE LOS HALLAZGOS DE LA INVESTIGACIÓN

OBJETIVOS

Después de esta sesión, usted debe ser capaz de:

1. *Idear una estrategia* para la difusión, la comunicación y el uso de los hallazgos de su investigación.
2. *Preparar la presentación* de sus hallazgos para los diversos sectores interesados.
3. *Preparar un plan de acción* para promover el uso de sus recomendaciones derivadas de la investigación.

- I. Introducción.
- II. Estrategia para la difusión, la comunicación y el uso de los hallazgos de la investigación.
- III. Presentación de los hallazgos de la investigación a los diversos sectores interesados.
- IV. Preparación de un plan de acción para promover el uso de las recomendaciones derivadas de la investigación.

I. INTRODUCCIÓN

Incluso los más grandes hallazgos de una investigación significan muy poco a menos que se difundan, se comuniquen y se usen eficazmente. Los beneficiarios de la investigación no son sólo los profesionales de la salud. Es necesario promover y aumentar la utilización de sus resultados de investigación entre todos los usuarios potenciales, que van desde los miembros de la comunidad hasta los organismos donantes. La probabilidad de que los hallazgos de la investigación se utilicen aumenta si se toman las siguientes medidas:

PASOS CLAVE

- (1) Conciba una estrategia de difusión sistemática y de comunicación, y utilícela para llegar a los diferentes grupos de usuarios potenciales
- (2) Presente los resultados de la investigación a todos los sectores interesados y reciba sus comentarios sobre las conclusiones y las recomendaciones
- (3) Elabore un plan de acción para promover la aplicación de las recomendaciones derivadas de su estudio

II. ESTRATEGIA PARA LA DIFUSIÓN Y LA COMUNICACIÓN DE LOS RESULTADOS DE LA INVESTIGACIÓN

El propósito de la ISS es proveer información útil a los funcionarios de *todos los niveles*, que les facilite solucionar problemas. Por lo tanto, un paso sumamente importante en el proceso de investigación sobre los sistemas de salud es la presentación de los resultados a todos los sectores interesados, de modo que pueda comentar con cada uno de ellos los hallazgos y las recomendaciones, así como la posibilidad de emprender acciones. La mera producción y difusión de un informe o un artículo de investigación no es, por lo general, suficiente.

Debe idear una estrategia de difusión y comunicación para promover el uso de los hallazgos de la investigación. Esta estrategia debe considerar los siguientes *elementos*:

1. Recapitulación del problema, los principales factores contribuyentes y acciones propuestas para resolverlo

Este resumen debe enfocarse en lo que se espera lograr con el estudio. Su contenido le ayuda a determinar a quiénes presentar los resultados, a fin de movilizarlos para la acción.

Por ejemplo, en un estudio de los factores que contribuyen a las frecuentes epidemias de cólera en el distrito de Kabwe, en Zambia, un importante factor contribuyente es, según se descubrió, la eliminación deficiente de los residuos metabólicos humanos, evidente cuando la cobertura de letrinas es de 10%. Los resultados deben presentarse por lo menos a los comités de salud y a los líderes de los poblados. Una recomendación importante es aumentar la cobertura de letrinas. Una posible acción es echar a andar un programa participativo de construcción de letrinas.

2. Identifique a los diferentes socios (grupos o instituciones) y su posible contribución para resolver el problema

Identifique claramente quienes conforman el grupo objetivo (blanco) de la acción, a la institución que vaya a ser responsable de la implementación, al personal que vaya a llevar a cabo las actividades del día a día, a los políticos que deban dar su visto bueno y que puedan proporcionar el apoyo financiero. Todos estos grupos e instituciones deben ser informados de los resultados de su estudio; deben tener la capacidad de responder a sus hallazgos y recomendaciones, y de comprometerse con algún tipo de acción.

Siempre es recomendable buscar apoyo “cerca a casa.” El equipo de investigación de Kabwe que llevó a cabo el estudio consistió, en parte, en los miembros el Equipo administrativo de salud del distrito (EASD). Seleccionaron a las siguientes instituciones y personas para pedirles sus comentarios:

- *Institución responsable:* Equipo administrativo de salud del distrito (EASD)
- *Grupo objetivo:* Líderes del poblado y comités de salud del poblado. En última instancia: jefes de familia
- *Apoyo técnico:* Programa comunitario de agua y sanidad del ministerio de salud; apoyo de ONG para el agua y actividades sanitarias
- *Apoyo político:* Funcionario médico del distrito (FMD), presidente del EASD
- *Apoyo financiero:* El EASD + ONG/donador centrado en el agua y la sanidad
- *Implementación día a día:* Los auxiliares en salud que apoyan a los jefes de familia

En este caso, se propone al equipo de gestión del *Sistema Integral Local de Salud* (EGSILOS) como el socio fundamental por parte del gobierno. El FMD

debe dar apoyo político para la recaudación de fondos adicionales, el funcionario de salud de distrito (FSD) –que fue parte del equipo de investigación– es responsable de la ejecución técnica. Tiene vínculos con el gerente de la Comunidad del Ministerio de Salud de Agua y Saneamiento, que debe proporcionar apoyo técnico, y con los auxiliares de salud en el ámbito que cae bajo su autoridad.

Los jefes de familia son los principales socios en la comunidad. Se les puede abordar mejor a través de los líderes del poblado y el comité de salud local (o comité de desarrollo local, si acaso el CSL no es lo suficientemente robusto). Los trabajadores de la salud del poblado, si los hubiera, también podrían ser socios importantes, así como los grupos de mujeres, puesto que las mujeres son un grupo importante de usuarias y encargadas de preservar la salud, y quizás estén escasamente representadas entre los jefes de familia. También los líderes religiosos pueden ser movilizados bien dispuestos. El apoyo de un miembro del consejo del distrito local puede ser igualmente útil.

Identificar a los socios financieros es crucial para aplicar las recomendaciones de la investigación. Los recursos deben buscarse, en primer lugar, en el presupuesto ordinario de las autoridades distritales y del Ministerio de Salud. Por lo general se necesitan fondos adicionales. En el distrito de Kabwe, se pudo identificar una ONG especializada en el campo del agua y la salubridad, dispuesta a brindar apoyo técnico y financiero.

En esta etapa también conviene considerar los *posibles obstáculos* a la acción propuesta. A veces hay grupos en la comunidad que están menos interesados en una actividad o incluso se oponen a ella. Debe darse una especial atención a estos grupos en el proceso de intercambio de opiniones respecto de los resultados de la investigación.

Por ejemplo, en un estudio sobre los factores que contribuyen a la notificación tardía para recibir tratamiento por parte de los enfermos de lepra en Aceh, Indonesia, se descubrió que muchos pacientes tomaban un tratamiento tradicional antes de acudir a un Centro de Salud. Ni los curanderos tradicionales (*dukun*) ni los miembros de la comunidad tenían la suficiente conciencia de la disponibilidad de un potente tratamiento breve. Cuando los resultados del estudio se volvieron a mostrar a los líderes de la comunidad y se comentaron las posibilidades de hacer una búsqueda intensiva de casos, se tuvo el cuidado de invitar también a los *dukun* locales.

3. Seleccione los canales de información adecuados

Una vez que se ha identificado a los socios, éstos deben ser notificados *personalmente*. En el caso del estudio de Kabwe, el EGSILOS se abordaría mejor en una *reunión normal*. Asimismo, sería fácil convocar una reunión de los líderes del

poblado, el Comité de desarrollo comunitario y los miembros de Comité de salud, los líderes religiosos y los representantes de los grupos de mujeres. Deben organizarse reuniones informativas personales entre los miembros del equipo de investigación y el administrador del Programa de agua y sanidad del Ministerio de Salud. La ONG debe abordarse por separado.

Ahora, o más adelante, cuando se planifique la acción, hay que pensar en los *medios de comunicación masiva* tales como los canales informativos: prensa, radio, televisión, carteles. Estos medios son eficaces para difundir información a una amplia gama de públicos. Los medios de comunicación masiva y los canales interpersonales pueden complementarse de manera fructífera. Identifique a los contactos clave de cada organización noticiosa; por ejemplo, a los periodistas que cubren el área de su investigación o al editor de las noticias. Puede usted abordarlos mediante una carta con información relevante sobre su proyecto. Puede utilizar las síntesis del estudio que ya tenga listas. Es conveniente abordar el problema de modo que la razón del estudio pueda ser plenamente entendida por personas ajenas al campo de la salud.

Además, otros investigadores pueden estar interesados en los resultados de la investigación y en los métodos empleados, sobre todo aquellos que se encuentran en los departamentos de salud comunitaria o de las Ciencias Sociales. Es apropiado publicar artículos en *revistas de investigación* y presentar los resultados de la investigación a los estudiantes y el personal interesado, incluso si dichas presentaciones no se relacionan directamente con la acción. También puede convenir presentar los resultados de los proyectos de ISS en “jornadas de investigación” organizadas para mejorar la comunicación entre un grupo selecto de funcionarios de la salud y los investigadores de la salud.

El *anexo 33.1* ofrece una visión general de una estrategia de difusión y comunicación de los hallazgos de la investigación con base en el ejemplo del estudio de la eliminación de los residuos metabólicos humanos en el distrito de Kabwe.

III. PRESENTACIÓN DE LOS RESULTADOS DE LA INVESTIGACIÓN A LOS DIFERENTES SECTORES INTERESADOS

Una vez identificadas las personas y las organizaciones con más probabilidades de asociarse para aplicar los resultados de la investigación –y los canales para abordarlas–, es hora de considerar la presentación en sí. Deben tomarse en cuenta una serie de puntos.

1. Cerciórese de que se asigne el tiempo suficiente a la presentación y los comentarios

- Prepare una presentación
- Programe el tiempo suficiente para su presentación (¡por lo menos unos 15 o 20 minutos!)
- Reserve el mayor tiempo posible para los comentarios a fin de obtener el mayor enriquecimiento posible en torno a sus hallazgos y sus recomendaciones
- Incluya los comentarios en los siguientes pasos (la acción)

2. Organice su presentación

Su presentación puede consistir en:

- Una breve introducción, que incluya el planteamiento del problema, los objetivos del estudio, las muestras y las herramientas de recolección empleadas.
- Los principales hallazgos enlistados en una secuencia lógica (por ejemplo, comenzando con una descripción del problema seguida de las principales variables que influyen en él).
- Las recomendaciones, en más o menos la misma secuencia.

De preferencia, separe las recomendaciones para los responsables de formular las políticas, para los funcionarios de salud, para el personal de la salud y para los miembros de la comunidad. Así le será más fácil hablar con cada grupo.

Por ejemplo, los comentarios sobre el estudio de Kabwe a los *miembros de la comunidad* deben insistir en la probabilidad de contraer el cólera a través del contacto con residuos metabólicos humanos, y en el papel de las letrinas en la prevención del cólera. Esta información debe venir como antecedente de los datos sobre la magnitud de la epidemia de cólera. La baja cobertura de letrinas y sus razones –arrojadas por el estudio–, tendrían que comentarse a profundidad para obtener sugerencias para emprender una intervención participativa de construcción de letrinas. Los comentarios para el *EGSILOS*, el *administrador del Programa de agua y sanidad* y la *ONG* deben destacar la asociación entre la prevalencia del cólera y la mala disposición de los residuos metabólicos en el distrito de Kabwe, las razones de la baja cobertura de letrinas y la reacción de los miembros de la comunidad ante la propuesta de emprender una intervención participativa para la construcción de letrinas.

Recuerde que su audiencia quiere saber básicamente: “¿Cómo podemos solucionar este problema?” Por lo tanto:

- Evite la jerga técnica
- No abrume a su público con datos estadísticos; sin embargo, puede presentar algunas tablas para apoyar sus principales conclusiones e ilustrar los problemas que identificó con algunas observaciones interesantes
- Sea *específico* en sus recomendaciones respecto de las acciones necesarias para resolver el problema; si su investigación indica que hay varias opciones viables, describa las alternativas y sus posibles ventajas y desventajas, y dé un amplio margen para los comentarios; tenga cuidado de no dar la impresión de estar “diciéndole a su público qué hacer”

3. Prepare un apoyo visual apropiado

Tenga suficientes copias de su “Resumen de los hallazgos y las recomendaciones” para todos los presentes. Si la presentación tiene lugar cuando usted ya terminó su informe, disponga de algunas copias del informe completo para quienes se muestren más interesados.

Prepare láminas de retroproyector, transparencias u hojas de rota-folio para destacar los puntos más importantes de su presentación (por ejemplo problema, principales objetivos, hallazgos más importantes y recomendaciones que impliquen la acción de aquellos a quienes se está usted dirigiendo).

4. Comente los hallazgos, la lógica y la factibilidad de las recomendaciones con los distintos grupos blanco

Infórmese para saber si están de acuerdo con sus conclusiones respecto de la naturaleza, la magnitud y las causas del problema, con base en su propia experiencia. Solicite información adicional sobre las preguntas que queden por contestar. Concéntrese en los comentarios sobre las recomendaciones que les conciernan, para obtener su opinión sobre su conveniencia y su factibilidad, y trate de obtener también su apoyo para las acciones que ellos mismos deban tomar. Infórmelos brevemente acerca de las recomendaciones hechas a otros sectores interesados, y de su reacción, y pídale sugerencias adicionales para la acción.

5. Designe a dos miembros del equipo para que lleven la minuta de cada sesión

Cerciórese de que se haga una minuta apropiada de los comentarios, especialmente en relación con las decisiones y las acciones de seguimiento acordadas.

Estas minutas deben distribuirse posteriormente entre todos los presentes, así como entre las personas clave que fueron invitadas, pero no pudieron asistir.

Trabajo en equipo, parte I

Prepare presentaciones para grupos selectos. Programe el tiempo suficiente para hacer la presentación –al menos medio día para esbozar el esquema de lo que va a presentar y cualesquiera notas que necesite, y para adaptar, enseguida, su “Resumen de los principales hallazgos y recomendaciones” para su distribución, de ser necesario. Reserve el resto del día para hacer correcciones, sacar copias del resumen y de los folletos, de haberlos, y preparar las láminas de retroproyector u hojas de rota-folio.

El líder del equipo o el miembro del equipo de investigación mejor dispuesto puede hacer toda la presentación, pero también se puede asignar a varios miembros del grupo para que cada uno se haga cargo de una parte de la presentación.

IV. PREPARACIÓN DE UN PLAN DE ACCIÓN

Cuando las presentaciones suscitan respuestas positivas, ha llegado el momento de *trazar un plan de acción* para el año entrante. Como parte de este plan, debe elaborarse el marco estratégico –comentado en la primera sección de este módulo–, incorporando los comentarios de la comunidad, de los funcionarios de la salud invitados y de los posibles donadores y ONG.

Si debe emprenderse una serie de acciones concretas para las que es necesario que diversas partes hagan un plan pormenorizado, puede *organizar un “taller de planificación de la acción.”* En este contexto los diversos grupos involucrados pueden trabajar juntos para estudiar de manera puntual los hallazgos y las recomendaciones, y elaborar un plan de acción. La presentación para los funcionarios y los líderes de la comunidad puede ampliarse e incluir esta fase de planificación activa, o puede programar un taller por separado, una vez que las personas clave responsables de la toma de decisiones hayan revisado los resultados del estudio y sus implicaciones.

La duración del taller puede variar desde un par horas hasta un par de días, dependiendo del tamaño del estudio y de la naturaleza de las acciones por planificar. Si opta por un formato más largo, considere una breve visita de campo a algunos de los sitios de la investigación antes de que el grupo comience la “fase de planificación de la acción” del taller.

Cuando se elaboren los planes, cerciórese de que los grupos de trabajo consideren las actividades y tareas que van a llevarse a cabo, quién se va a hacer responsable, cuándo se van a realizar y qué recursos se van a necesitar.

Como ejemplo de qué tanto se puede desarrollar un plan de acción, tomemos la recomendación de comenzar un proceso de *intervención participativa de construcción de letrinas* del estudio de Kabwe.

Suponiendo que todos los sectores abordados (miembros de la comunidad, el EGSILOS, el Programa comunitario de agua y sanidad, las ONG) hubieran aceptado participar, habría que especificar los pormenores de esta participación. Podría proponerse que el equipo de investigación, apoyado por el EGSILOS, organizara los siguientes talleres de una jornada:

- (1) Un taller con representantes del Programa comunitario de agua y sanidad, de la ONG, el funcionario de salud del distrito y de algunos auxiliares del sector salud, para comentar los pormenores técnicos y el costo de los diferentes tipos de letrinas.
- (2) Un taller con diferentes líderes de la comunidad, (+ representantes de las ADC, trabajadores de la salud del poblado, clubes de mujeres, y líderes religiosos) sobre los deseos y las posibilidades con respecto a la construcción de letrinas en la comunidad, seguida por reuniones similares en todos los barrios del distrito.
- (3) Un taller con todos los socios sobre el objetivo final, los tipos de letrinas y la estimación aproximada del costo, la contribución material de todos los socios y demás acciones de apoyo necesarias.

Las reuniones deben cubrir los siguientes puntos de acción:

1. En primer lugar, debe especificarse el objetivo de la intervención. Por ejemplo: un aumento de la cobertura de letrinas en el distrito de Kabwe de 10% a 80% de los hogares, en un lapso de tres años.
2. Enseguida debe especificarse la *contribución material de todos los socios*. Los jefes de familia deben estar dispuestos a contribuir con mano de obra y parte del material de construcción y del equipo; el equipo de gestión de la salud del distrito debe estar dispuesto a incorporar el programa de construcción de letrinas en sus actividades de salud ambiental, proporcionando mano de obra (auxiliares de salud, funcionario de salud de distrito) y el transporte; se le puede pedir a la ONG interesada que apoye a los jefes de familia con material de construcción adicional y que apoye al EGSILOS pagando el transporte adicional y el costo de la capacitación de los auxiliares y los trabajadores de la salud del poblado. Por otra parte, se le puede pedir a la ONG que dé incentivos anuales a las 25 letrinas mejor construidas y a las 25 con mejor mantenimiento en el distrito.
3. Deben especificarse, finalmente, las demás acciones de apoyo institucional en todos los planos. En el plano de la comunidad, el CSL debe estar

dispuesto a asumir la responsabilidad de movilizar los grupos familiares para construir letrinas y darles mantenimiento (por ejemplo, en las reuniones del poblado, a través de los clubes de mujeres, etc.). El club de jóvenes de la iglesia puede ayudar a los miembros del CSL, por ejemplo, con representaciones teatrales sobre la higiene y el papel de las letrinas en el control del cólera y otras enfermedades. El CSL también puede apoyar a los trabajadores de la salud, a quienes también se capacitaría para que puedan brindar asesoría técnica en la construcción de letrinas, y supervisar y controlar su construcción en el poblado. En el plano del EGSÍLOS, el funcionario responsable del medio ambiente del distrito debe hacer un plan de trabajo pormenorizado y un presupuesto para tres años, que incluya todas las actividades de movilización, capacitación, construcción y monitoreo, seguimiento y evaluación. Para el plano de los hogares, el *Departamento del agua y la sanidad del Ministerio de Salud* y la *ONG* de apoyo deben estar dispuestos a ayudar en la elaboración de hojas de instrucciones sencillas para las construcción, material de capacitación para los auxiliares de salud ambiental y material de IEC para las escuelas (siempre y cuando las escuelas estén dispuestas a participar).

Se puede abordar a los *medios de comunicación masiva* para proporcionar mayor información acerca de la intervención participativa planificada para construir letrinas tan pronto como el primer poblado haya comenzado esta labor. Se podría pensar en un artículo periodístico con fotografías y una transmisión en la radio o la televisión sobre el problema (el cólera, la baja cobertura de letrinas), la intervención planificada y los primeros resultados.

El *anexo 33.2* presenta un resumen esquemático del plan de acción propuesto.

Trabajo en equipo, parte II

1. Prepare un plan de acción para implementar las recomendaciones de su estudio
2. Si el plan para aplicar los resultados debe involucrar la participación de diversas personas u organizaciones, considere si conviene organizar un “taller de planificación de la acción”
3. Si así fuera, haga un plan específico; envíe las notas de las discusiones previas al respecto, formule los objetivos de las reuniones y acuerde una fecha con los socios invitados
4. Durante los talleres, precise tanto como sea posible la naturaleza y el calendario de las contribuciones y haga un plan de trabajo viable para las diferentes actividades. El EGSÍLOS, en particular la persona responsable de la ejecución del plan de acción, debe preparar, después de los talleres, un plan de trabajo más preciso para un periodo dado (1-3 años), que combine los planes de trabajo tentativos para las distintas actividades.

Módulo 33 DIFUSIÓN, COMUNICACIÓN Y USO DE LOS HALLAZGOS DE LA INVESTIGACIÓN

Tiempo estimado y recursos didácticos

30 minutos

Introducción y discusión

Introducción y discusión

Es importante presentar este módulo antes de que los participantes inicien el resumen de sus hallazgos y sus recomendaciones.

Nota:

Subraye que el PLAN DE ACCIÓN debe considerar cómo movilizar a los sectores interesados cuya participación es crítica para resolver o atenuar el problema investigado.

Trabajo en equipo

Al presentar los resultados de la investigación es recomendable que *los participantes distribuyan el resumen –con las recomendaciones completas–, a todos los participantes e invitados de honor* para fomentar su participación en el debate. Los pormenores de los objetivos específicos del estudio, el método de investigación y los hallazgos (por ejemplo, algunas tablas o gráficos cruciales) pueden presentarse en hojas de rota-folio o en láminas de retroproyector.

Anexo 33.1 Estrategia para la difusión y la comunicación de los hallazgos de la investigación, para promover su uso (estudio de la eliminación de los residuos metabólicos humanos en el distrito de Kabwe, Zambia)

Problema principal	Principal acción propuesta	Socios potenciales	Canales informativos	Instrumentos y principales puntos de discusión	Resultados esperados
Epidemia de cólera	Programa participativo de construcción de letrinas	<i>Comunidad:</i>	Reuniones del poblado	Resumen del informe Retroproyector Carteles Explicación oral Debate	Acuerdo para
		– Líderes del poblado			– la participación en la construcción de letrinas, su mantenimiento y su uso
		– Comité de salud del poblado			
Deficiente eliminación de los residuos metabólicos humanos (10% de cobertura de letrinas)		– Trabajadores de la salud			
		– Grupos de mujeres			
		– Líderes religiosos			
		<i>Ministerio de salud</i>	Reuniones regulares	– Las condiciones de poca higiene causan la diseminación del cólera	– Apoyo político y técnico
		– Equipo de gestión de la salud del distrito			
		– Comité de la salud del poblado	Resúmenes personales	– Necesidad de aumentar la cobertura de letrinas	– Recursos necesarios (mano de obra, algo de dinero y equipo)
Falta de conocimiento, falta de apoyo		– Trabajadores de la salud			
		– Auxiliares de la salud ambiental	Reunión con el funcionario del medio ambiente del distrito		
		<i>ONG interesadas en el agua y la sanidad</i>	Resumen personal	– Posibilidades de participación en el programa de construcción de letrinas	Apoyo técnico y financiero adicional
		<i>Medios de comunicación masiva</i>	Resumen personal		– Artículo periodístico sobre los hallazgos de la investigación y programa de construcción de letrinas planeado

Anexo 33.2 Estrategia del plan de acción para una intervención participativa para la construcción de letrinas, distrito de Kabwe, Zambia)

<i>Objetivo acordado</i>	<i>Socios</i>	<i>Contribución material</i>	<i>Acción de apoyo institucional adicional</i>	<i>Resultados esperados</i>	<i>Resultados</i>
Aumento de la cobertura de letrinas de 10% a 80% en tres años	<i>Comunidad:</i>	<i>Jefes de familia</i>	– Reuniones del poblado	– Aumento de la cobertura de letrinas de 10% a 80% en tres años	Una importante reducción en la gravedad y la frecuencia del cólera (en última instancia: su eliminación)
	– Líderes del poblado	– Trabajo		– Aumento del mantenimiento y el uso de las letrinas	
	– Comité de salud del poblado	– Parte del material (local) y el equipo		– Mejor higiene y conducta sanitaria	
	– Trabajadores de la salud				
	– Grupos de mujeres				
	– Líderes religiosos				
	<i>Ministerio de salud</i>	<i>Ministerio de salud /DHMT</i>			
	– Equipo de gestión de la salud del distrito	– Mano de obra (env asistentes de la salud + DHO)	– Reuniones regulares		
	– Comité de la salud del poblado	– Transporte (parcial)	– Resúmenes personales		
	– Trabajadores de la salud	– Asistentes de salud ambiental			
	– Asistentes de la salud ambiental		– Reunión con el funcionario del medio ambiente del distrito		
	<i>ONG interesadas en el agua y la sanidad</i>	<i>ONG</i>	<i>ONG</i>		
		– Transporte adicional y costo de capacitación	– Asesoría para redactar instrucciones de construcción sencillas y material de capacitación para trabajadores de la salud, EHA y IEC		
		– Material de construcción adicional			
		– Incentivo anual a las 25 letrinas mejor construidas y a las 25 mejor conservadas			
	Medios de comunicación masiva		Medios de comunicación masiva		
			– Material para las escuelas		
			– Artículos periodísticos		
			– Transmisiones por radio o actividades planificadas y realizadas		

Este espacio en blanco intencionalmente dejado de la página

ACERCA DE LOS AUTORES

Corlien M. Varkevisser, MA, PhD, MPH, es socióloga-antropóloga médica de profesión, especialista en salud pública. Como miembro del Royal Tropical Institute de Ámsterdam, y presidente fundador de la Primary Health Care (PHC) Unit, ha acumulado una amplia experiencia en la investigación sobre sistemas de salud y administración PHC en el subsahara africano. Fue una de los cofundadores del proyecto Joint HSR (OMS/Netherlands Ministry for Development Cooperation/Royal Tropical Institute) para Sudáfrica y fue asignada a la Oficina subregional de la OMS en Harare como directora del Proyecto Joint HSR –desde su fundación en abril de 1987 hasta 1992. Desde entonces es directora del curso MPH en el RTI y profesora de ISS en la Facultad de Ciencias Políticas y Sociales de la Universidad de Ámsterdam.

Indra Pathmanathan, MMBS, MPH, es médica especializada en salud pública; como directora del programa de ISS del Ministerio de Salud en Malasia desde sus inicios, ha sido responsable del desarrollo y la instrumentación de diversas estrategias para el ISS, que se han replicado en otros países. Éstas incluyen programas de capacitación en ISS y Control de calidad para la toma de decisiones en ministerios, para médicos y para el personal de los equipos de salud distritales, hospitales y universidades. Fue miembro del Grupo Asesor en ISS, OMS-Ginebra y participó en el consejo editorial de BRIDGE. Los últimos diez años ha sido consultora del Banco Mundial en el campo de la salud.

Ann Brownlee, MA, PhD, es socióloga y médica especialista en ISS, planificación y evaluación, así como en aspectos transculturales del cuidado de la salud. Durante varios años fue Coordinadora de investigación y evaluación del Project for Strengthening Health Delivery Systems en África occidental y África central. Trabajó muy de cerca con la Oficina Regional de la OMS en África y con colegas de esa nación y otras regiones para desarrollar un programa de capacitación en ISS y publicar el *Curso de Capacitación en ISS*, que fue precursor de este volumen. Trabaja actualmente como consultora en salud internacional para la OMS, el CIID, y Wellstart, entre otros. Es catedrática de la Universidad de California en San Diego.